

A comparison of nonlinear population Monte Carlo and particle Markov chain Monte Carlo algorithms for Bayesian inference in stochastic kinetic models

Eugenia Koblents · Joaquín Míguez

Received: date / Accepted: date

Abstract In this paper we address the problem of Monte Carlo approximation of posterior probability distributions in stochastic kinetic models (SKMs). SKMs are multivariate Markov jump processes that model the interactions among species in biochemical systems according to a set of uncertain parameters. Markov chain Monte Carlo (MCMC) methods have been typically preferred for this Bayesian inference problem. Specifically, the particle MCMC (pMCMC) method has been recently shown to be an effective, while computationally demanding, method applicable to this problem. Within the pMCMC framework, importance sampling (IS) has been used only as the basis of the sequential Monte Carlo (SMC) approximation of the acceptance ratio in the Metropolis-Hastings kernel. However, the recently proposed nonlinear population Monte Carlo (NPMC) algorithm, based on an iterative IS scheme, has also been shown to be effective as a Bayesian inference tool for low dimensional (predator-prey) SKMs. In this paper, we provide an extensive performance comparison of pMCMC versus NPMC, when applied to the challenging prokaryotic autoregulatory network. We show how the NPMC method can greatly outperform the pMCMC algorithm in this scenario, with an overall moderate computational effort. We complement the numerical comparison of the two techniques with an asymp-

totic convergence analysis of the nonlinear IS scheme at the core of the proposed method when the importance weights can only be computed approximately.

Keywords Nonlinear population Monte Carlo · particle Markov chain Monte Carlo · sequential Monte Carlo · stochastic kinetic models

1 Introduction

Stochastic kinetic models (SKMs) are multivariate systems that model molecular interactions among species in biological and chemical problems, according to a set of unknown rate parameters (Wilkinson, 2011b). The aim of this paper is the approximation of the posterior distribution of the rate parameters and the populations of all species, provided a set of discrete, noisy observations is available. This inference problem has been traditionally addressed using Markov chain Monte Carlo (MCMC) schemes (Boys et al, 2008; Milner et al, 2013; Wilkinson, 2011a,b). In (Golightly and Wilkinson, 2011) a particle MCMC (pMCMC) method (Andrieu et al, 2010) has been successfully applied to this problem. The pMCMC technique relies on a sequential Monte Carlo (SMC) approximation of the posterior distribution of the populations to compute the Metropolis-Hastings (MH) acceptance ratio.

However, MCMC methods in general, and pMCMC in particular, suffer from a number of problems. The convergence of the Markov chain is hard to assess and the final set of samples presents correlations which can greatly reduce its efficiency. Besides, MCMC methods do not (easily) allow for parallel implementations and turn out to be computationally intensive. To reduce the complexity of the existing MCMC methods

E. K. acknowledges the support of *Ministerio de Educación* of Spain (*Programa de Formación de Profesorado Universitario*, ref. AP2008-00469). This work has been partially supported by *Ministerio de Economía y Competitividad* of Spain (program Consolider-Ingenio 2010 CSD2008-00010 COMONSENS and project COMPREHENSION TEC2012-38883-C02-01).

E. Koblents and J. Míguez
Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Madrid, Spain
E-mail: ekoblents,jmiguez@tsc.uc3m.es

when applied to SKMs, a diffusion approximation of the underlying stochastic process is usually applied (Golightly and Wilkinson, 2005). The parameters of the MCMC proposal are also hard to choose and determine the performance of the algorithm.

An appealing alternative to the widely established MCMC methods is the population Monte Carlo (PMC) algorithm (Cappé et al, 2004). PMC is an iterative importance sampling (IS) scheme that yields a discrete approximation of a target probability distribution. The PMC algorithm has important advantages with respect to MCMC techniques. It provides independent samples and asymptotically unbiased estimates at all iterations, which avoids the need of a convergence period. Additionally, PMC may be easily parallelized.

On the other hand, the main weakness of IS and PMC is their low efficiency in high dimensional problems, due to the well known degeneracy problem (Bengtsson et al, 2008). The recently proposed nonlinear PMC (NPMC) scheme (Koblents and Míguez, 2013b) mitigates this difficulty by computing nonlinear transformations of the importance weights (IW), in order to smooth their variations and avoid degeneracy. In (Koblents and Míguez, 2013b) a simple convergence analysis of nonlinear IS (NIS) is provided, for two types of nonlinear transformations, *tempering* and *clipping*. Similarly to the pMCMC method in (Golightly and Wilkinson, 2011), the NPMC method resorts to an SMC approximation of the posterior distribution of the populations to compute, in our case, the IWs.

In (Koblents and Míguez, 2013a,c) the nonlinear version of IS and PMC is combined with the popular mixture-PMC (MPMC) method of (Cappé et al, 2008), which allows to approximate arbitrary high-dimensional target distributions by means of mixtures of Gaussian or t-Student distributions. The original MPMC algorithm of (Cappé et al, 2008) has been applied to cosmological inference problems and compared to an MCMC method in (Wraith, 2009) (and (Kilbinger, 2010)), and has been shown to provide similar precision results with a lower computation load than its MCMC counterpart. The MPMC scheme is the basis of the tool CosmoPMC (Kilbinger, 2012) for the estimation of cosmological parameters, as an alternative to the MCMC package, CosmoMC, (Lewis and Bridle, 2002) <http://cosmologist.info/cosmomc>.

In this paper we apply the NPMC method to the estimation of both the parameters and the unobserved populations in SKMs. We present numerical results to compare the performance of the state-of-art pMCMC and the proposed NPMC, when applied to the challenging prokaryotic model in two scenarios of different

dimension and with two different observation models. We show that the NPMC method outperforms the pMCMC method for the same computational cost.

As a complement to the numerical comparison, we introduce new asymptotic convergence results for the NIS scheme that accounts for the use of SMC to approximate the IWs. The analysis in this paper considerably extends the preliminary results in (Koblents and Míguez, 2013b). In particular, we prove that approximate integrals computed via NIS converge almost surely (as the number of samples increases) and explicit convergence rates are given.

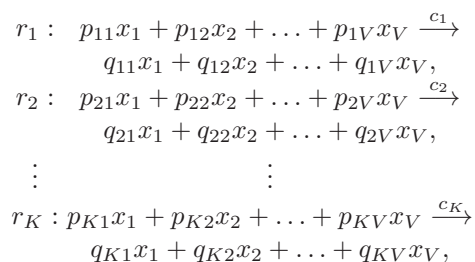
The rest of the paper is organized as follows. In Section 2 we present an introduction to the basics of SKMs and the usual solutions to this Bayesian inference problem. In Sections 3 and 4 we describe the pMCMC and NPMC methods, respectively, when applied to the approximation of posterior distributions in SKMs. In Section 5 we numerically compare the performance of pMCMC and NPMC schemes when applied to a prokaryotic autoregulatory model, with different simulation settings. Section 6 is devoted to the convergence analysis of the NIS method. Finally, Section 7 is devoted to the conclusions.

2 Bayesian inference for stochastic kinetic models

2.1 Stochastic kinetic models

A SKM is a multivariate continuous-time jump process modeling the interactions among molecules, or species, that take place in chemical reaction networks of biochemical and cellular systems (Wilkinson, 2011b).

Consider a biochemical reaction network that describes the time evolution of the population of V species x_1, \dots, x_V related by means of K reactions r_1, \dots, r_K



where p_{kv} and q_{kv} , $k = 1, \dots, K$, $v = 1, \dots, V$, denote the reactant and the product coefficients, respectively; and $c_k > 0$, $k = 1, \dots, K$, are the random constant rate parameters. A matrix \mathbf{P} of size $K \times V$ contains the reactant coefficients p_{kv} and, similarly, \mathbf{Q} contains the product coefficients q_{kv} . The stoichiometry matrix of size $V \times K$ is defined as $\mathbf{S} = (\mathbf{Q} - \mathbf{P})^\top$. The vector $\mathbf{c} = [c_1, \dots, c_K]^\top$ contains the rate parameters.

Let $x_v(t)$, $v = 1, \dots, V$, denote the nonnegative, integer population of species x_v at time t , and let $\mathbf{x}(t) = [x_1(t), \dots, x_V(t)]^\top$ denote the state of the system at this time instant. Let $\mathbf{x}_n = [x_{1,n}, \dots, x_{V,n}]^\top$ denote the state of the system at discrete time instants $t = n\Delta$, $n = 1, \dots, N$, i.e., $x_{v,n} = x_v(n\Delta)$ where Δ denotes a time-discretization period. We denote by \mathbf{x} the $VN \times 1$ vector containing the population of each species at N consecutive discrete time instants, i.e., $\mathbf{x} = [\mathbf{x}_1^\top, \dots, \mathbf{x}_N^\top]^\top$.

The k -th reaction takes place stochastically according to its instantaneous rate or hazard function

$$h_k(t) = c_k \prod_{v=1}^V \binom{x_v(t)}{p_{kv}}, \quad k = 1, \dots, K,$$

where the product of binomial coefficients represents the number of combinations in which the k -th reaction can occur, as a function of the population of each reactant species x_v . We additionally define the vector $\mathbf{h}(t) = [h_1(t), \dots, h_K(t)]^\top$. The waiting time to the next reaction is exponentially distributed with parameter $h_0(t) = \sum_{k=1}^K h_k(t)$, and the probability of each reaction type is given by $h_k(t)/h_0(t)$.

2.2 Bayesian inference for SKMs

We consider the log-transformed rate parameters $\boldsymbol{\theta} = [\theta_1, \dots, \theta_K]^\top$, where $\theta_k = \log(c_k)$, $k = 1, \dots, K$, with prior pdf $p(\boldsymbol{\theta})$. The prior pdf of the initial population vector \mathbf{x}_0 is denoted by $p(\mathbf{x}_0)$. We assume that a linear combination of the populations of a subset of species is observed at discrete time instants corrupted by Gaussian noise, i.e.,

$$\mathbf{y}_n = \mathbf{M}\mathbf{x}_n + \mathbf{w}_n, \quad n = 1, \dots, N, \quad (1)$$

where \mathbf{M} is the observation matrix with dimensions $D \times V$ and $\mathbf{w}_n \sim \mathcal{N}_D(\mathbf{w}_n; \mathbf{0}, \sigma^2 \mathbf{I})$ is a multivariate Gaussian noise component. We denote the complete observation vector with dimension $DN \times 1$ as $\mathbf{y} = [\mathbf{y}_1^\top, \dots, \mathbf{y}_N^\top]^\top$.

The dynamical behavior of an arbitrary SKM may be described in terms of the following set of equations¹

$$\begin{cases} \boldsymbol{\theta} \sim p(\boldsymbol{\theta}) & (\text{parameters prior}), \\ \mathbf{x}_0 \sim p(\mathbf{x}_0) & (\text{populations prior}), \\ \mathbf{x}_n \sim p(\mathbf{x}_n | \mathbf{x}_{n-1}, \boldsymbol{\theta}) & (\text{transition equation}), \\ \mathbf{y}_n \sim p(\mathbf{y}_n | \mathbf{x}_n) & (\text{observation equation}), \end{cases}$$

¹ For simplicity of notation, in this section we use p to denote the pdfs in the model. We write conditional pdfs as $p(\mathbf{y} | \mathbf{x})$, and joint densities as $p(\boldsymbol{\theta}) = p(\theta_1, \dots, \theta_K)$. This is an argument-wise notation, hence $p(\theta_1)$ denotes the distribution of θ_1 , possibly different from $p(\theta_2)$.

where $p(\mathbf{x}_n | \mathbf{x}_{n-1}, \boldsymbol{\theta})$ and $p(\mathbf{y}_n | \mathbf{x}_n)$ denote the transition pdf and the likelihood function, respectively. The Gillespie algorithm (Gillespie, 1977) allows to perform exact forward simulations of arbitrary SKMs, drawing samples from the transition densities $p(\mathbf{x}_n | \mathbf{x}_{n-1}, \boldsymbol{\theta})$, $n = 1, \dots, N$, given a set of log-rate parameters $\boldsymbol{\theta}$ and an initial population \mathbf{x}_0 .

In this paper, we aim to obtain a Monte Carlo approximation of the full joint posterior distribution of the log-rate parameters $\boldsymbol{\theta}$ and the populations \mathbf{x} , with density

$$p(\boldsymbol{\theta}, \mathbf{x} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{x}) p(\mathbf{x} | \mathbf{x}_0, \boldsymbol{\theta}) p(\mathbf{x}_0) p(\boldsymbol{\theta}), \quad (2)$$

given the prior distributions $p(\boldsymbol{\theta})$ and $p(\mathbf{x}_0)$, the transition pdf $p(\mathbf{x} | \mathbf{x}_0, \boldsymbol{\theta}) = \prod_{n=1}^N p(\mathbf{x}_n | \mathbf{x}_{n-1}, \boldsymbol{\theta})$ and the likelihood function $p(\mathbf{y} | \mathbf{x}) = \prod_{n=1}^N p(\mathbf{y}_n | \mathbf{x}_n)$ constructed from equation (1).

We are also interested in computing approximations of the posterior marginals of the rate parameters $p(\boldsymbol{\theta} | \mathbf{y}) = \int p(\boldsymbol{\theta}, \mathbf{x} | \mathbf{y}) d\mathbf{x}$ and the species populations $p(\mathbf{x} | \mathbf{y}) = \int p(\boldsymbol{\theta}, \mathbf{x} | \mathbf{y}) d\boldsymbol{\theta}$ as well as their moments (e.g., the posterior mean), which are of the form

$$E_{p(\boldsymbol{\theta} | \mathbf{y})}[f(\boldsymbol{\theta})] = \int f(\boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}, \text{ and}$$

$$E_{p(\mathbf{x} | \mathbf{y})}[f(\mathbf{x})] = \int f(\mathbf{x}) p(\mathbf{x} | \mathbf{y}) d\mathbf{x}, \text{ respectively,}$$

where f is a real, integrable function.

Bayesian inference based on exact stochastic simulations from $p(\mathbf{x}_n | \mathbf{x}_{n-1}, \boldsymbol{\theta})$ generated via the Gillespie algorithm often becomes practically intractable even for models of modest complexity (Golightly and Wilkinson, 2005). Thus, it is very common to resort to a continuous approximation of the underlying stochastic process, which is known as the diffusion approximation. The diffusion process that most closely matches the dynamics of the associated Markov jump process, over an infinitesimal time interval $(t, t + dt]$, is given by a stochastic differential equation known as the chemical Langevin equation (CLE) (Wilkinson, 2011b) (pag 230)

$$d\mathbf{x}(t) = \mathbf{S} \mathbf{h}(t) dt + \sqrt{\mathbf{S} \text{diag}\{\mathbf{h}(t)\} \mathbf{S}^\top} d\mathbf{w}(t),$$

driven by the $V \times 1$ dimensional Wiener process $\mathbf{w}(t)$. However, this approximation is known to be poor in low concentration scenarios, and thus should be avoided for models involving species with a very low population. Alternatively, in (Milner et al, 2013) the authors propose a solution based on a moment closure approximation of the stochastic process.

This inference problem has been traditionally addressed using MCMC methods, and IS based schemes

have been avoided due to their inefficiency in high dimensional spaces (Wilkinson, 2011b). In (Boys et al, 2008) various MCMC algorithms are evaluated in data-poor scenarios. In (Golightly and Wilkinson, 2011) a likelihood-free pMCMC scheme (Andrieu et al, 2010) is applied to this problem. This method is, to the best of our knowledge, the most powerful, yet computationally expensive, method provided so far for this kind of applications.

In (Koblenets and Míguez, 2013b) a NPMC scheme is proposed for the approximation of the marginal posterior pdf $p(\boldsymbol{\theta}|\mathbf{y})$, which is computationally competitive, since it requires the processing of a low number of samples of $\boldsymbol{\theta}$ to obtain the approximation of the posterior. The performance of the NPMC method is tested in a simple SKM known as predator-prey model (Volterra, 1926), providing excellent results with a low computational cost.

In this paper we compare the performances of the pMCMC and the NPMC methods in the approximation of the full joint posterior $p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{y})$ in equation (2), which allows to perform Bayesian inference for the rate parameters $\boldsymbol{\theta}$ and the full sample path \mathbf{x} , including unobserved components.

3 Particle MCMC for SKMs

The particle marginal Metropolis-Hastings (PMMH) algorithm is a pMCMC method originally proposed in (Andrieu et al, 2010) for Monte Carlo sampling from the full posterior distribution $p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{y})$. The PMMH scheme suggests a proposal mechanism of the form $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})\hat{p}^J(\mathbf{x}^*|\mathbf{y}, \boldsymbol{\theta}^*)$. A new candidate in the parameter space, $\boldsymbol{\theta}^*$, is drawn from an arbitrary proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$, while the new candidate in the variable space, \mathbf{x}^* , is generated using an approximation of the posterior marginal $p(\mathbf{x}^*|\mathbf{y}, \boldsymbol{\theta}^*)$ constructed by means of an SMC algorithm (i.e., a particle filter) with J particles and denoted $\hat{p}^J(\mathbf{x}^*|\mathbf{y}, \boldsymbol{\theta}^*)$. The probability of accepting the proposed pair $(\boldsymbol{\theta}^*, \mathbf{x}^*)$ is

$$\min \left\{ 1, \frac{\hat{p}^J(\mathbf{y}|\boldsymbol{\theta}^*)p(\boldsymbol{\theta})}{\hat{p}^J(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}^*)} \times \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta})} \right\},$$

where $\hat{p}^J(\mathbf{y}|\boldsymbol{\theta}^*)$ is an unbiased approximation of the marginal likelihood of $\boldsymbol{\theta}^*$ (i.e., $p(\mathbf{y}|\boldsymbol{\theta}^*)$), computed, again, by way of a particle filter with J particles. The PMMH algorithm is reproduced in Table 1, and the SMC approximations of $p(\mathbf{y}|\boldsymbol{\theta}^*)$ and $p(\mathbf{x}^*|\mathbf{y}, \boldsymbol{\theta}^*)$ are described in Appendix A. Full details can be found in (Andrieu et al, 2010). Note that the forward simulation of the stochastic process in the particle filter may be performed exactly with the Gillespie algorithm, or using a diffusion approximation.

Table 1 Particle MCMC algorithm targeting $p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{y})$ (Andrieu et al, 2010).

Initialization ($i = 0$):

1. Sample $\boldsymbol{\theta}^{(0)} \sim p(\boldsymbol{\theta})$ and
2. run a SMC scheme targeting $p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}^{(0)})$. Draw $\mathbf{x}^{(0)} \sim \hat{p}^J(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}^{(0)})$ from the SMC approximation and let $\hat{p}^J(\mathbf{y}|\boldsymbol{\theta}^{(0)})$ denote the marginal likelihood estimate.

Iteration ($i = 1, \dots, I$):

1. Sample $\boldsymbol{\theta}^* \sim q(\cdot|\boldsymbol{\theta}^{(i-1)})$ and
2. run a SMC scheme targeting $p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}^*)$. Draw $\mathbf{x}^* \sim \hat{p}^J(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}^*)$, let $\hat{p}^J(\mathbf{y}|\boldsymbol{\theta}^*)$ denote the marginal likelihood estimate, and
3. with probability

$$\min \left\{ 1, \frac{\hat{p}^J(\mathbf{y}|\boldsymbol{\theta}^*)p(\boldsymbol{\theta})}{\hat{p}^J(\mathbf{y}|\boldsymbol{\theta}^{(i-1)})p(\boldsymbol{\theta}^{(i-1)})} \times \frac{q(\boldsymbol{\theta}^{(i-1)}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i-1)})} \right\}$$

accept the move setting $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^*$, $\mathbf{x}^{(i)} = \mathbf{x}^*$ and $\hat{p}^J(\mathbf{y}|\boldsymbol{\theta}^{(i)}) = \hat{p}^J(\mathbf{y}|\boldsymbol{\theta}^*)$. Otherwise store the current values $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i-1)}$, $\mathbf{x}^{(i)} = \mathbf{x}^{(i-1)}$ and $\hat{p}^J(\mathbf{y}|\boldsymbol{\theta}^{(i)}) = \hat{p}^J(\mathbf{y}|\boldsymbol{\theta}^{(i-1)})$.

In (Golightly and Wilkinson, 2011) the proposal is selected as a Gaussian random walk $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}) = \mathcal{N}_K(\boldsymbol{\theta}^*; \boldsymbol{\theta}, \gamma^2)$, whose variance γ^2 has to be tuned and partly determines the performance of the algorithm.

After removing the initial burn-in samples and thinning the output, we obtain a Markov chain $\{\boldsymbol{\theta}^{(i)}, \mathbf{x}^{(i)}\}_{i=1}^M$ with M correlated samples. Then, we may construct a sample approximation of the marginal posterior distributions of the parameters $\boldsymbol{\theta}$ and the populations \mathbf{x} , as

$$\hat{p}^M(d\boldsymbol{\theta}|\mathbf{y}) = \frac{1}{M} \sum_{i=1}^M \delta_{\boldsymbol{\theta}^{(i)}}(d\boldsymbol{\theta}) \quad \text{and}$$

$$\hat{p}^M(d\mathbf{x}|\mathbf{y}) = \frac{1}{M} \sum_{i=1}^M \delta_{\mathbf{x}^{(i)}}(d\mathbf{x}),$$

respectively, where $\delta_{\boldsymbol{\theta}^{(i)}}$ and $\delta_{\mathbf{x}^{(i)}}$ denote the unit delta measure centered at $\boldsymbol{\theta}^{(i)}$ and $\mathbf{x}^{(i)}$, respectively. The approximation of the full joint posterior is of the form

$$\hat{p}^M(d\boldsymbol{\theta}, d\mathbf{x}|\mathbf{y}) = \frac{1}{M} \sum_{i=1}^M \delta_{(\boldsymbol{\theta}^{(i)}, \mathbf{x}^{(i)})}(d\boldsymbol{\theta}, d\mathbf{x}).$$

4 Nonlinear PMC for SKMs

The PMC method (Cappé et al, 2004) is an iterative IS scheme that generates a sequence of proposal pdf's $q_\ell(\cdot)$, $\ell = 1, \dots, L$, that approximate a target pdf π along the iterations. In (Koblenets and Míguez, 2013b) the NPMC scheme is proposed, which introduces nonlinearly transformed IWs (TIWs) in order to mitigate

the numerical problems caused by degeneracy in the proposal update scheme.

We first consider as a target density the marginal posterior pdf of the parameters θ given the observation vector \mathbf{y} , i.e., $\pi(\theta) = p(\theta|\mathbf{y})$. As in (Koblents and Míguez, 2013b), we construct the proposal pdf $q_\ell(\theta)$, $\ell = 2, \dots, L$, as a Gaussian approximation of the target pdf obtained at the previous iteration $\ell - 1$, whose mean and covariance parameters are selected to match the moments of the previous sample set. The NPMC algorithm is displayed in Table 2. Details and some simple convergence results can be found in (Koblents and Míguez, 2013b).

Table 2 Nonlinear PMC targeting $\pi(\theta) = p(\theta|\mathbf{y})$.

Iteration ($\ell = 1, \dots, L$):

1. Draw a set of M samples $\{\theta_\ell^{(i)}\}_{i=1}^M$ from the proposal density $q_\ell(\theta)$:
 - at iteration $\ell = 1$, let $q_1(\theta) = p(\theta)$.
 - at iterations $\ell = 2, \dots, L$ the proposal $q_\ell(\theta)$ is the Gaussian approximation of $p(\theta|\mathbf{y})$ obtained at iteration $\ell - 1$.
2. For $i = 1, \dots, M$, run a SMC scheme with J particles targeting $p(\mathbf{x}|\mathbf{y}, \theta_\ell^{(i)})$ and compute the marginal likelihood estimate $\hat{p}_\ell^J(\mathbf{y}|\theta_\ell^{(i)})$.
3. For $i = 1, \dots, M$, compute the unnormalized IWs

$$w_\ell^{(i)*} \propto \frac{\hat{p}_\ell^J(\mathbf{y}|\theta_\ell^{(i)})p(\theta_\ell^{(i)})}{q_\ell(\theta_\ell^{(i)})}.$$

4. For $i = 1, \dots, M$, compute normalized TIWs, $\bar{w}_\ell^{(i)}$, by *clipping* the original IWs as

$$\bar{w}_\ell^{(i)*} = \min(w_\ell^{(i)*}, \mathcal{T}_\ell^{M_T}), \quad \bar{w}_\ell^{(i)} = \bar{w}_\ell^{(i)*} / \sum_{j=1}^M \bar{w}_\ell^{(j)*},$$

where the threshold value $\mathcal{T}_\ell^{M_T}$ denotes the M_T -th highest unnormalized IW $w_\ell^{(i)*}$, with $1 < M_T < M$.

5. Resample to obtain an unweighted set $\{\tilde{\theta}_\ell^{(i)}\}_{i=1}^M$: for $i, j = 1, \dots, M$, let $\tilde{\theta}_\ell^{(i)} = \theta_\ell^{(j)}$ with probability $\bar{w}_\ell^{(j)}$.
6. Construct a Gaussian approximation $q_{\ell+1}(\theta) = \mathcal{N}(\theta; \mu_\ell, \Sigma_\ell)$ of the posterior $p(\theta|\mathbf{y})$, where the mean vector and covariance matrix are computed as

$$\mu_\ell = \frac{1}{M} \sum_{i=1}^M \tilde{\theta}_\ell^{(i)} \quad \text{and} \quad \Sigma_\ell = \frac{1}{M} \sum_{i=1}^M (\tilde{\theta}_\ell^{(i)} - \mu_\ell)(\tilde{\theta}_\ell^{(i)} - \mu_\ell)^\top. \quad (3)$$

Equivalently to the pMCMC algorithm, in the NPMC implementation the densities $p(\mathbf{x}|\mathbf{y}, \theta)$ and $p(\mathbf{y}|\theta)$ required in steps 2 and 3 are replaced by their SMC approximations, which are given in Appendix A. The NPMC method may also use either exact or ap-

proximate samples of the stochastic process, depending on the computational capabilities.

For the *clipping* procedure performed in step 4 we consider, at each iteration ℓ , a permutation i_1, \dots, i_M of the indices in $\{1, \dots, M\}$ such that $w_\ell^{(i_1)*} \geq \dots \geq w_\ell^{(i_M)*}$ and choose a *clipping* parameter $M_T < M$. We select a threshold value $\mathcal{T}_\ell^M = w_\ell^{(i_{M_T})*}$ and apply *clipping* to the largest IWs $w_\ell^{(i_k)*} \geq \mathcal{T}_\ell^M$, $k = 1, \dots, M_T - 1$. This transformation leads to M_T flat TIWs in the region of interest of θ , allowing for a robust update of the proposal. The performance of the algorithm is robust to the selection of the *clipping* parameter M_T (Koblents and Míguez, 2013b). For simplicity, step 5 performs multinomial resampling.

At each iteration of the NPMC algorithm we may construct a discrete approximation of the posterior pdf $p(\theta|\mathbf{y})$, based on the set of samples and TIWs, as

$$\hat{p}_\ell^M(d\theta|\mathbf{y}) = \sum_{i=1}^M \bar{w}_\ell^{(i)} \delta_{\theta_\ell^{(i)}}(d\theta).$$

The choice of a Gaussian approximation of the proposal $q_{\ell+1}(\theta)$ in step 6 is arbitrary (and done for simplicity here). Any other family of pdfs can be used without modifying the rest of the algorithm (Koblents and Míguez, 2013a,c).

4.1 NPMC targeting $p(\theta, \mathbf{x}|\mathbf{y})$

The NPMC method proposed in (Koblents and Míguez, 2013b) may be readily applied to the approximation of the full joint posterior $p(\theta, \mathbf{x}|\mathbf{y})$, in a manner equivalent to the pMCMC algorithm. We consider a sampling mechanism of the form $q(\theta)\hat{p}^J(\mathbf{x}|\mathbf{y}, \theta)$, where samples $\theta^{(i)}$ are again generated from the latest proposal $q(\theta)$ and $\mathbf{x}^{(i)}$ are drawn from the SMC approximation $\hat{p}^J(\mathbf{x}|\mathbf{y}, \theta^{(i)})$ obtained via particle filtering (the iteration index has been omitted for simplicity). Then, the standard, unnormalized IW associated to the pair $(\theta^{(i)}, \mathbf{x}^{(i)})$ is computed as

$$w^{(i)*} = \frac{\hat{p}^J(\theta^{(i)}, \mathbf{x}^{(i)}|\mathbf{y})}{q(\theta^{(i)})\hat{p}^J(\mathbf{x}^{(i)}|\mathbf{y}, \theta^{(i)})} \propto \frac{\hat{p}^J(\mathbf{x}^{(i)}, \mathbf{y}|\theta^{(i)})p(\theta^{(i)})}{q(\theta^{(i)})\hat{p}^J(\mathbf{x}^{(i)}|\mathbf{y}, \theta^{(i)})} \propto \frac{\hat{p}^J(\mathbf{y}|\theta^{(i)})p(\theta^{(i)})}{q(\theta^{(i)})}$$

and is independent of \mathbf{x} . This reveals that, when samples $\mathbf{x}_\ell^{(i)}$ are drawn from $\hat{p}^J(d\mathbf{x}|\mathbf{y}, \theta)$ the algorithm yields a discrete approximation of the posterior distribution of the unobserved populations \mathbf{x} constructed as

$$\hat{p}_\ell^M(d\mathbf{x}|\mathbf{y}) = \sum_{i=1}^M \bar{w}_\ell^{(i)} \delta_{\mathbf{x}_\ell^{(i)}}(d\mathbf{x}).$$

Even though the proposed NPMC and the pMCMC require very similar computations for each pair of samples of $\{\boldsymbol{\theta}, \mathbf{x}\}$, and thus have an equivalent computational cost, the NPMC has a set of important advantages with respect to its MCMC counterpart. PMC methods in general can be more easily parallelized, drastically reducing their execution time. Additionally, they provide independent sets of samples at all iterations, and do not require a burn-in period. On the other hand, the nonlinearity applied in the NPMC mitigates weight degeneracy, which is the main problem arising in conventional IS based methods, dramatically increasing its efficiency in high-dimensional problems. As a consequence, we claim that the number of samples (and thus, the computational complexity) required by the NPMC can be significantly lower than that of pMCMC. Finally, contrary to pMCMC, which requires a careful choice of the proposal tuning parameter, the proposed method does not require the precise fitting of any parameters.

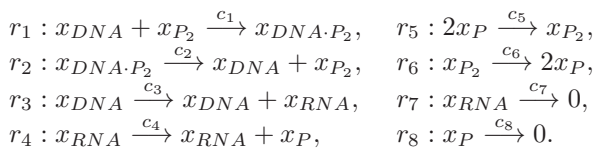
An extensive numerical comparison of pMCMC versus NPMC for the prokaryotic autoregulatory network is presented in Section 5.

5 Example: Prokaryotic autoregulatory model

In this section, we compare the performance of the pMCMC and the NPMC methods when applied to the problem of approximating the posterior distributions of the log-rate parameters $p(\boldsymbol{\theta}|\mathbf{y})$ and the populations $p(\mathbf{x}|\mathbf{y})$ in a simplified prokaryotic autoregulatory model, given some observed data \mathbf{y} . This problem has been introduced in (Golightly and Wilkinson, 2005), and further analyzed in (Golightly and Wilkinson, 2011; Wilkinson, 2011b). This prokaryotic model is minimal in terms of the level of details included and offers a simplistic view of the mechanisms involved in gene autoregulation. However, it contains many of the interesting features of an auto-regulatory feedback network and does provide sufficient detail to capture the network dynamics.

5.1 Prokaryotic autoregulatory model

The prokaryotic autoregulatory model is a SKM that involves $V = 5$ chemical species and $K = 8$ reaction equations, r_1, \dots, r_K , given by (Golightly and Wilkinson, 2005)



We construct the V -dimensional vector containing the population of each species at time instant t as $\mathbf{x}(t) = [x_{RNA}(t), x_P(t), x_{P_2}(t), x_{DNA \cdot P_2}(t), x_{DNA}(t)]^\top$. Thus, we obtain a stoichiometry matrix of the form

$$\mathbf{S} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 2 & 0 & -1 \\ -1 & 1 & 0 & 0 & 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and the hazard vector is given by

$$\mathbf{h}(t) = [c_1 x_{DNA} x_{P_2}, c_2 x_{DNA \cdot P_2}, c_3 x_{DNA}, c_4 x_{RNA}, c_5 \frac{x_P(x_P - 1)}{2}, c_6 x_{P_2}, c_7 x_{RNA}, c_8 x_P]^\top, \quad (4)$$

where the time dependance of the population of each species is omitted for notational simplicity.

This model involves a conservation law given by the relation $x_{DNA \cdot P_2} + x_{DNA} = C$, where C is the number of copies of this gene in the genome. We could use this relation to remove $x_{DNA \cdot P_2}$ from the model, replacing any occurrences of the latter in the hazard function with $C - x_{DNA}$, but in this paper we abide by the notation in equation (4). Further details of this model can be found in (Wilkinson, 2011b).

5.2 Simulation setup

We have selected most of the simulation parameters following (Golightly and Wilkinson, 2011). The true vector of rate parameters which we aim to estimate has been set to

$$\mathbf{c} = [0.1, 0.7, 0.35, 0.2, 0.1, 0.9, 0.3, 0.1]^\top,$$

which yields log-transformed rate parameters

$$\boldsymbol{\theta} = -[2.30, 0.36, 1.05, 1.61, 2.30, 0.10, 1.20, 2.30]^\top.$$

The initial populations and the conservation constant have been set to $\mathbf{x}_0 = [x_1(0), \dots, x_V(0)]^\top = [8, 8, 8, 5, 5]^\top$ and $C = 10$, respectively. The time discretization period is $\Delta = 1$ and the Gaussian noise variance is $\sigma^2 = 4$ (assumed to be known). In all the simulations in this paper we have performed exact sampling from the stochastic model with the Gillespie algorithm to obtain the likelihood approximation via particle filtering. The number of particles for the SMC approximation $\hat{p}^J(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$, has been set to $J = 100$ for all the simulations.

Independent uniform priors $\mathcal{U}(\theta_k; -7, 2)$ are taken for each $\theta_k = \log(c_k)$. Opposite to (Golightly and Wilkinson, 2011), the initial populations \mathbf{x}_0 are assumed unknown

for the inference algorithm and we consider independent Poisson priors $p(x_v(0)) = \mathcal{P}(x_v(0); \lambda_v)$, with λ_v parameters set to the true initial populations, that is, $\lambda_v = x_v(0)$, $v = 1, \dots, V$.

We consider two different observation scenarios. In the complete observation (CO) scenario we assume that all species x_v , $v = 1, \dots, V$, are observed at regular time intervals of length Δ and corrupted by Gaussian noise. Thus, the observation matrix is of the form $\mathbf{M} = \mathbf{I}_V$ and the observations are given by

$$\mathbf{y}_n = \mathbf{x}_n + \mathbf{w}_n, \quad n = 1, \dots, N.$$

In the CO case the complete vector of observations $\mathbf{y} = [\mathbf{y}_1^\top, \dots, \mathbf{y}_N^\top]^\top$ has dimension $VN \times 1$.

In the partial observation scenario (PO) only a linear combination of the proteins $x_P + 2x_{P_2}$ is observed, also contaminated by Gaussian noise, i.e., the observation matrix is given by $\mathbf{M} = [0, 1, 2, 0, 0]$ (with dimension $1 \times V$) and the observations are generated as

$$y_n = x_{2,n} + 2x_{3,n} + w_n, \quad \text{where } w_n \sim \mathcal{N}_1(w_n; 0, \sigma^2).$$

In the PO case, a vector of scalar observations with dimension $N \times 1$ is constructed as $\mathbf{y} = [y_1, \dots, y_N]^\top$.

5.3 Performance evaluation

To evaluate the performance of the pMCMC and the NPMC methods we compute, in all the simulation runs, the mean square error (MSE) attained by the sample set that approximates the marginal posterior of $\boldsymbol{\theta}$, generated by both schemes.

For the pMCMC method, we compute the MSE of each parameter θ_k based on the M -size final output (after removing the burn-in period and thinning), as

$$MSE_k = \frac{1}{M} \sum_{i=1}^M (\theta_k^{(i)} - \theta_k)^2, \quad k \in \{1, \dots, K\}.$$

For the NPMC, we compute the MSE associated to each parameter θ_k , $k = 1, \dots, K$, based on the unweighted sample set at the ℓ -th iteration $\{\tilde{\theta}_\ell^{(i)}\}_{i=1}^M$, $\ell = 1, \dots, L$, as

$$MSE_{\ell,k} = \frac{1}{M} \sum_{i=1}^M (\tilde{\theta}_{\ell,k}^{(i)} - \theta_k)^2 = (\mu_{\ell,k} - \theta_k)^2 + \sigma_{\ell,k}^2,$$

where $\mu_{\ell,k}$ is the k -th component of the mean vector $\boldsymbol{\mu}_\ell$ and the variance term $\sigma_{\ell,k}^2$ is the (k, k) component of matrix $\boldsymbol{\Sigma}_\ell$.

However, the MSE cannot be computed in real problems, where the true parameters θ_k are unknown. To monitor the stability and the efficiency of the two sampling schemes based on the generated sample alone, we

resort to the so called normalized effective sample size (NESS), which is often defined differently for MCMC and IS schemes (Robert and Casella, 2004).

In the MCMC literature, the NESS gives the relative size of an i.i.d. (independent and identically distributed) sample with the same variance as the current sample and thus indicates the loss in efficiency due to the use of a Markov chain (Robert and Casella, 2004). For pMCMC we compute the NESS from the final chain (after removing the burn-in period and thinning) as

$$M^{neff} = \frac{1}{1 + 2 \sum_{j=1}^{\infty} \hat{\rho}(j)},$$

where $\hat{\rho}(j) = \text{corr}(\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(j)})$ is the average autocorrelation function (ACF) at lag j . For the computation of the NESS, we truncate j when $\hat{\rho}(j) < 0.1$.

For IS methods, the NESS may be interpreted as the relative size of a sample generated from the target distribution with the same variance as the current sample. Even when high values of the NESS do not guarantee a low approximation error, the NESS is often used as an indicator of the numerical stability of the algorithm (Doucet et al, 2000). It cannot be evaluated exactly but we may compute an approximation of the NESS at each iteration of the NPMC scheme based on the set of TIWs as

$$M_\ell^{neff} = \frac{1}{M \sum_{i=1}^M (\bar{w}_\ell^{(i)})^2}, \quad \ell = 1, \dots, L.$$

5.4 Simulation results

We consider two simulation scenarios in which a different number of parameters is estimated.

5.4.1 Estimation of a single rate parameter θ_1

In this section we present numerical results regarding the approximation of the posterior distribution $p(\theta_1, \mathbf{x} | \boldsymbol{\theta}_{\setminus 1}, \mathbf{y})$ of a single rate parameter $\theta_1 = \log c_1$ and the populations \mathbf{x} , when the rest of parameters $\boldsymbol{\theta}_{\setminus 1} = [\theta_2, \dots, \theta_K]^\top$, are assumed to be known.

We compare the pMCMC and the NPMC methods in this simple scenario in order to illustrate the optimal performance of both schemes, in the CO and PO scenarios. These simulations show the degradation of the approximations when the amount of observations reduces.

We have performed $P = 100$ independent simulation runs of the pMCMC and the NPMC schemes in the CO and the PO scenarios, with different (independent) population and observation vectors in each simulation.

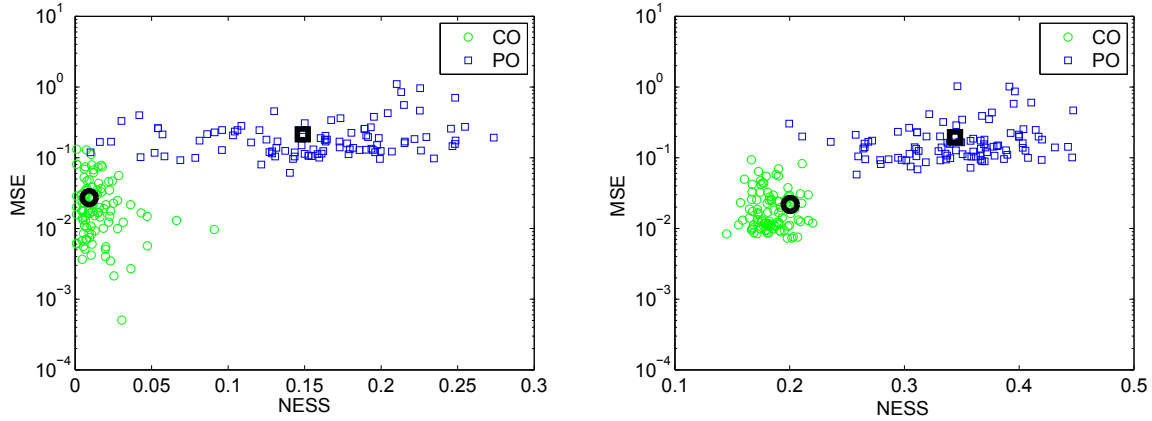


Fig. 1 Performance of the pMCMC (*left*) and the NPMC (*right*) methods for the estimation of a unique rate parameter θ_1 : MSE (in logarithmic scale) obtained from the final output versus the NESS for each simulation run in the CO and the PO scenario. The big circles and squares represent simulation runs with a final mean MSE close to the global average

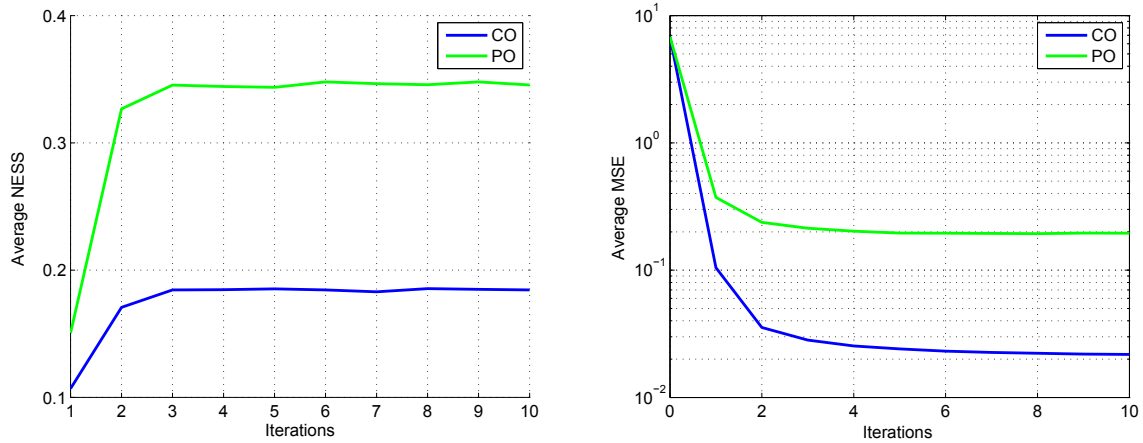


Fig. 2 Evolution along the iterations of the NPMC algorithm of the average NESS (*left*) and MSE (*right*) in the CO and PO scenarios, estimating a single parameter θ_1 .

Both in the CO and the PO cases, the same true population trajectories $\mathbf{x}^{(p)}$, $p = 1, \dots, P$, were used, but the observations in the CO scenario, $\mathbf{y}_{CO}^{(p)}$, and in the PO scenario, $\mathbf{y}_{PO}^{(p)}$, differ. The number of observation times has been set to $N = 100$.

As a proposal pdf $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$ in the pMCMC scheme we consider a Gaussian random walk update with variance $\gamma^2 = 1$, which to the best results in the simulations. A total number of $I = 10^4$ iterations has been run in each simulation. A final sample of size $M = 10^3$ has been obtained from each Markov chain by discarding a burn-in period of 10^3 samples and thinning the output by a factor of 9.

In the NPMC scheme, the number of iterations has been set to $L = 10$, the number of samples per iteration is $M = 10^3$ and the *clipping* parameter is $M_T = 100$. In this way, the computational effort of the two methods

is approximately the same, as they both generate 10^4 samples in the space of $\boldsymbol{\theta}$.

In Figure 1 the final MSE obtained by the pMCMC (*left*) and the NPMC (*right*) algorithms for each simulation run is depicted versus the final NESS, in the CO and the PO scenarios. Note that the NESS is computed differently for pMCMC and NPMC. It can be observed that both algorithms perform similarly in this case, with an equivalent computational cost. Both algorithms attain on average lower MSE values in the CO scenario, as expected. However, the NESS also takes lower values in the CO case, which indicates a worse mixing of the Markov chains in the pMCMC algorithm and also higher degeneracy in the NPMC algorithm.

In Figure 2 the evolution of the MSE (*right*) and the NESS (*left*) along the iterations of the NPMC algorithm is represented, for the CO and the PO scenarios. It can be observed that both measures attain a steady

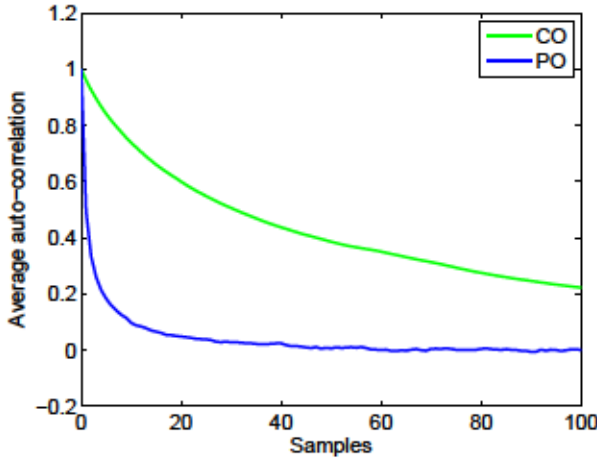


Fig. 3 Average ACF based on the final sample of size $M = 10^3$ of the pMCMC scheme in the CO and the PO scenarios, averaged over $P = 100$ simulation runs

value by the 5-th iteration, both in the CO and the PO case, which suggest that actually less iterations are sufficient for this problem. Again, we observe that in the CO scenario both the NESS and the MSE reach lower values.

Figure 3 plots the average ACF of the final pMCMC sample, after removing the burn-in period and thinning the Markov chain by a factor of 9. Particularly high correlations are present in the CO case, leading to a poor NESS. Related to the ACF, the average sample acceptance probability in the pMCMC scheme in the PO scenario is 0.091, while in the CO scenario it is only 0.0034. Which means that 910 samples are accepted out of $I = 10^4$ in the CO case and only 34 in the CO case.

In Figure 4 the final pdf estimates $\hat{p}(\theta_1|\theta_{\setminus 1}, \mathbf{y})$ of the average simulation runs represented as big circles and crosses in Figure 1 are represented in the CO and the PO scenario, for the pMCMC and the NPMC schemes. For the pMCMC method we have built a Gaussian approximation of the posterior density $p(\theta_1|\theta_{\setminus 1}, \mathbf{y})$ based on the final MCMC sample $\{\theta_1^{(i)}\}_{i=1}^M$. For the NPMC method, this approximation corresponds to the proposal pdf for the next iteration $L+1$, i.e., $\hat{p}(\theta_1|\theta_{\setminus 1}, \mathbf{y}) = q_{L+1}(\theta_1) = \mathcal{N}(\theta_1; \mu_{L,1}, \sigma_{L,1}^2)$, where the mean and variance terms $\mu_{L,1}$ and $\sigma_{L,1}^2$ are computed as in Eq. (3). It can be observed in Figure 4 that very similar results are obtained by both algorithms in this scenario. The final MSE values obtained by the pMCMC and the NPMC methods, averaged over $P = 100$ simulation runs, are shown in Table 3, together with the MSE corresponding to the prior distribution.

Figure 5 depicts the posterior mean of the populations, $\hat{\mathbf{x}} = E_{\hat{p}(\mathbf{x}|\mathbf{y})}[\mathbf{x}]$, obtained with pMCMC (left) as $\hat{\mathbf{x}} = \frac{1}{M} \sum_{i=1}^M \mathbf{x}^{(i)}$ and with NPMC (right) as $\hat{\mathbf{x}} =$

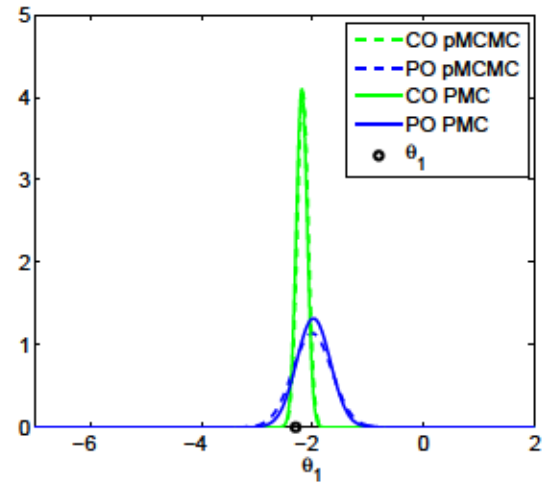


Fig. 4 Marginal posterior pdf estimates $\hat{p}(\theta_1, |\theta_{\setminus 1}, \mathbf{y})$ of an average simulation run, for pMCMC and NPMC in the CO and PO scenarios. The true value θ_1 is also shown

Table 3 Final mean and standard deviation (std) values of the MSE for θ_1 in the CO and PO scenarios, for pMCMC and NPMC. The prior values are included for comparison

		mean MSE	std MSE
Prior		6.789	0
PO	pMCMC	0.215	0.171
	NPMC	0.195	0.170
CO	pMCMC	0.027	0.026
	NPMC	0.022	0.016

$\sum_{i=1}^M \bar{w}_L^{(i)} \mathbf{x}_L^{(i)}$ in the PO scenario. The results correspond to the particular simulation runs (different for pMCMC and NPMC) identified with big squares in Figure 1 and whose posterior approximations, $\hat{p}(\theta_1|\theta_{\setminus 1}, \mathbf{y})$, are shown in Figure 4. It can be observed that, in the PO scenario, the tendency of the population of all the species is reasonably identified, even though only a linear combination of the proteins is observed. In the CO scenario the populations of all species are accurately estimated and are not shown for conciseness. Note that the populations of all species are very low, which suggests that the diffusion approximation may perform poorly in this scenario.

The results presented in this section reveal a very similar performance of the two methods in this simple scenario. Also in terms of computational complexity pMCMC and NPMC perform very similarly. The execution time per 10^3 samples (one NPMC iteration and 10^3 pMCMC iterations) for the pMCMC scheme is 312 seconds, while for NPMC it is 325 seconds, both in the CO and in the PO cases, on a 3-GHz Intel Core 2 Duo CPU, with 2 GB of RAM. The stochastic forward simulation of the prokaryotic model with the Gillespie

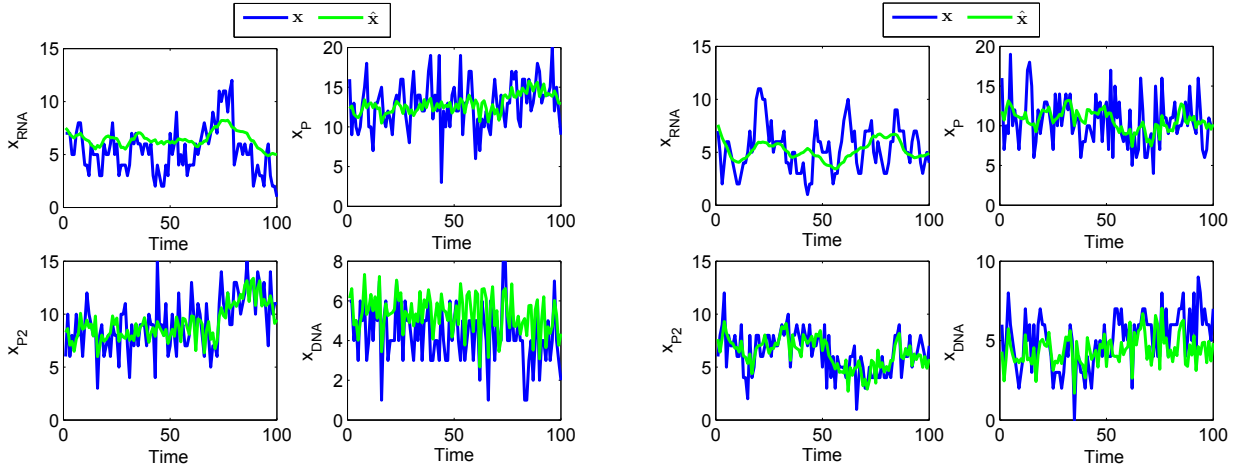


Fig. 5 Posterior mean, $\hat{\mathbf{x}} = E_{\hat{p}(\mathbf{x}|\mathbf{y})}[\mathbf{x}]$, of the populations obtained in a single simulation run of pMCMC (*left*) and NPMC (*right*) in the PO scenario (only a linear combination of the proteins is observed, corrupted by noise)

algorithm has been implemented in C, and the rest of the code in Matlab R2007b.

However, the pMCMC method provides a set of highly correlated samples, specially in the CO scenario, and requires the setting of the proposal variance γ^2 as well as the burn-in period length and the thinning parameter, which may not be straightforward and determines the performance of the algorithm. On the contrary, the NPMC scheme provides uncorrelated sets of samples at each iteration, and does not require the precise fitting of any parameters. Additionally, the computer simulations suggest that the convergence of the NPMC algorithm may be assessed observing the evolution of the NESS, which usually reaches a steady value simultaneously with the MSE.

5.4.2 Estimation of all the parameters θ_k , $k = 1, \dots, K$

In this section we present simulation results to evaluate the performance of the pMCMC and the NPMC schemes in the approximation of the posterior distribution of the rate parameters and the populations of all species, $p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{y})$, assuming that all the rate parameters are unknown, again in the CO and the PO scenarios.

In this case, $N = 200$ observation times are assumed for all the simulations. Again, $P = 100$ independent simulation runs of each algorithm have been performed. The NPMC scheme has been run for $L = 15$ iterations, with $M = 10^3$ samples per iteration and *clipping* parameter $M_T = 100$. The pMCMC scheme has been run with $I = 15 \times 10^3$ iterations in each simulation run, a burn-in period of 10^3 iterations and thinning the output by a factor of 14. With this setup the computational effort is approximately the same in the two schemes.

In Figure 6 the MSE (in logarithmic scale), averaged over the parameters θ_k , attained by the pMCMC (*left*) and the NPMC (*right*) algorithms is represented versus the NESS, in the CO and PO scenarios. Simulation runs which attained a final MSE close to the global average value are indicated with big circles (CO) and squares (PO) on both plots. It can be observed that the pMCMC method performs similarly in both scenarios, in terms of MSE and NESS, yielding poor results in both cases. On the contrary, the NPMC method provides significantly better *MSE* results in the CO scenario, where a larger amount of information is available. The NPMC method does not present degradation due to the high degeneracy occurring in the CO scenario.

Figure 7 depicts the evolution along the iterations of the NESS (*left*) and the MSE (*right*) averaged over $P = 100$ independent simulation runs for the NPMC algorithm. Both indices converge to a steady value in a low number of iterations also in this complex scenario. As expected, a significantly higher final MSE is attained in the extremely data poor PO scenario.

In Figure 8 (*left*) the average ACF attained by the pMCMC in the CO and the PO cases is represented. Even after thinning the output, the sample correlation is extremely high in both scenarios, which leads to a very low NESS. The acceptance rate is also very low and very long chains are required to obtain reasonable results. In the PO scenario 43.69 samples are accepted on average in a simulation run of $I = 15 \times 10^3$ samples (acceptance rate 0.0029). In the CO case, only 23.07 samples are accepted on average (rate 0.0015).

Figure 8 (*right*) depicts the final Markov chain provided by the pMCMC method (after removing the burn-in period and thinning the output) in the average simulation run represented with a big square in Figure

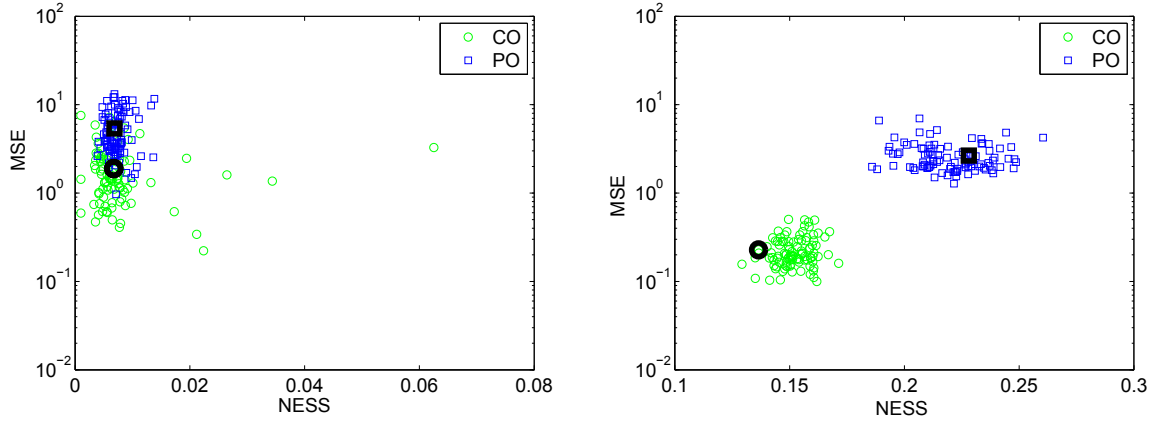


Fig. 6 Performance of the pMCMC (*left*) and the NPMC (*right*) methods for the estimation of the whole set of rate parameters θ : MSE (in logarithmic scale) versus the final NESS, for each simulation run in the CO and the PO scenario. The big circles and squares represent simulation runs with a final mean MSE close to the global average

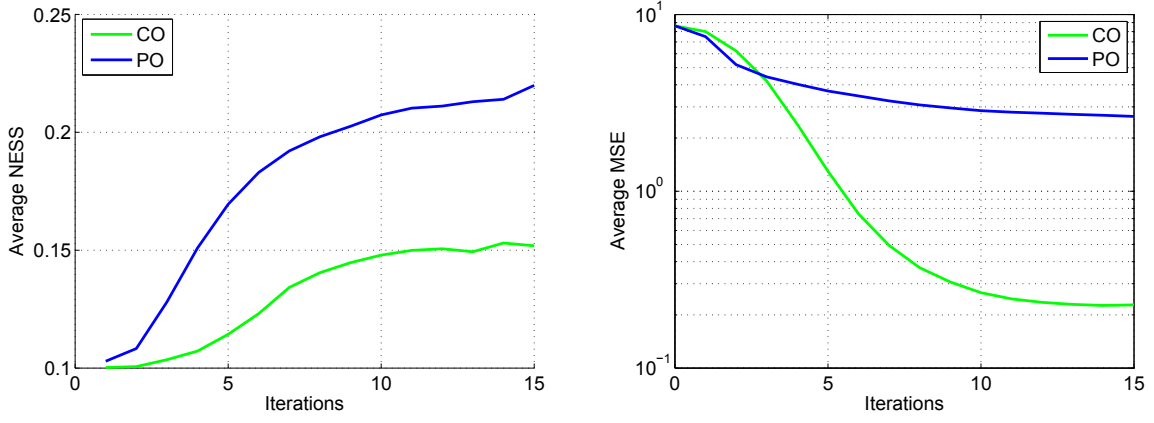


Fig. 7 Evolution along the NPMC iterations of the average NESS (*left*) and MSE (*right*) in the CO and the PO scenario.

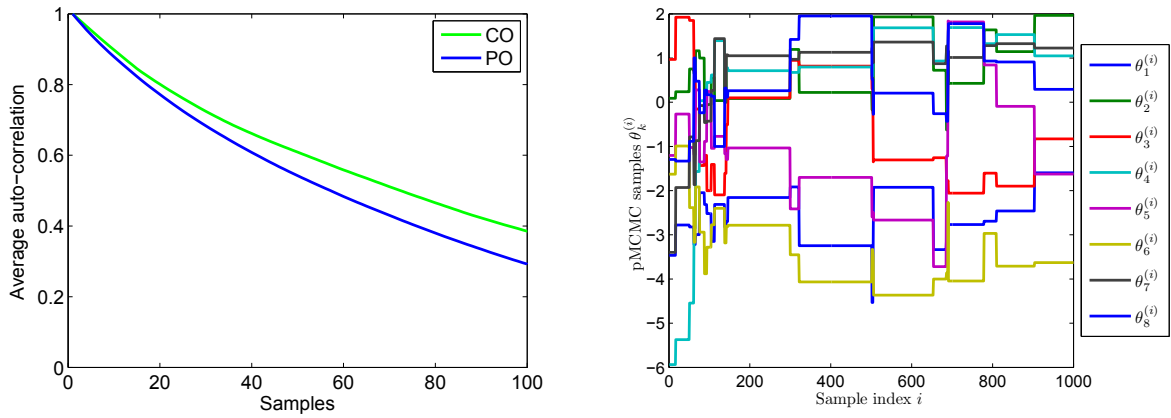


Fig. 8 *Left*: Auto-correlations based on the final sample of size 10^3 of the pMCMC scheme in the CO and the PO scenarios, averaged over $P = 100$ simulation runs. *Right*: Markov chain provided by the pMCMC method in the PO scenario, corresponding to the average simulation run depicted with a big square in Figure 6 (*left*).

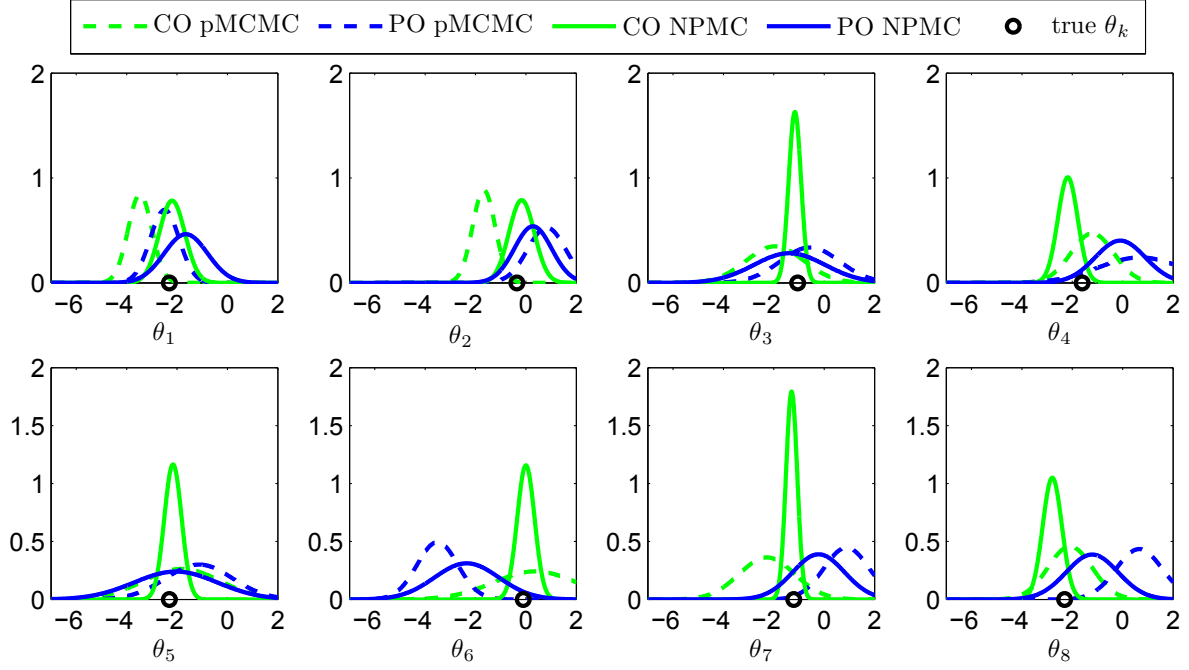


Fig. 9 Marginal posterior pdf approximations of each parameter $\hat{p}(\theta_k|\mathbf{y})$, $k = 1, \dots, K$, attained in an average simulation run by the pMCMC and the NPMC, in the CO and in the PO case.

Fig. 10 Final MSE for the parameters θ_k , $k = 1, \dots, K$ in the CO and PO experiments, averaged over the simulation runs. The last two columns corresponds to the mean and standard deviation (std) values of the global MSE (averaged over the parameters). The prior values are included for comparison

		θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	mean MSE	std MSE
Prior		6.789	11.344	8.853	7.543	6.789	12.484	8.430	6.789	8.628	0
PO	pMCMC	3.412	3.319	5.543	3.200	7.059	8.929	6.799	4.371	5.329	2.926
	NPMC	1.246	1.011	2.214	1.490	4.073	7.015	2.311	1.856	2.652	1.020
CO	pMCMC	2.899	2.958	1.676	1.572	1.604	1.547	1.573	1.468	1.912	1.476
	NPMC	0.305	0.302	0.162	0.167	0.280	0.280	0.156	0.168	0.228	0.091

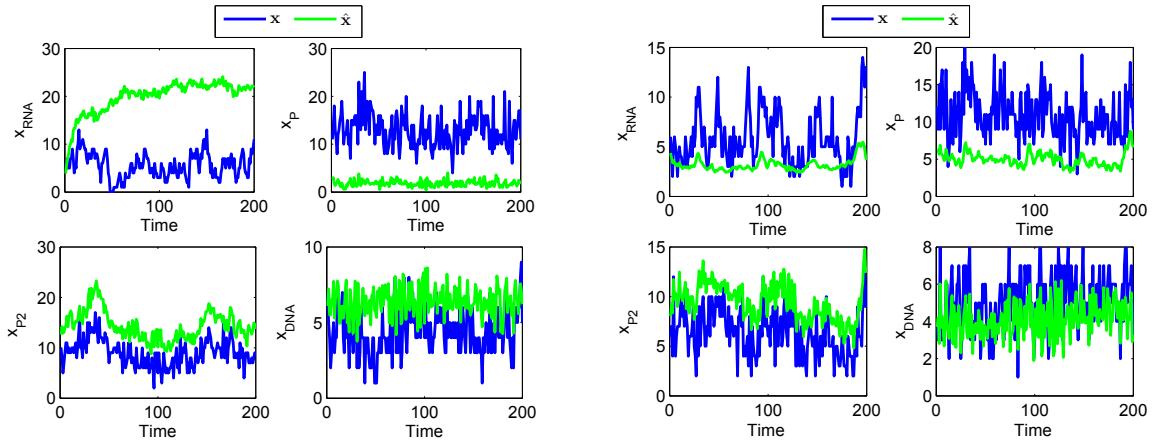


Fig. 11 Posterior mean $\hat{\mathbf{x}} = E_{p(\mathbf{x}|\mathbf{y})}[\mathbf{x}]$ of the populations of all species obtained in the average simulation run of the pMCMC (left) and the NPMC (right) schemes, in the PO scenario.

6 (*left*). It can be observed that the mixing of the chain is very poor, with a total number of accepted samples of 46 (close to the average). Many other simulations, both in the PO and the CO scenarios, provide even lower number of accepted samples, and thus, very inconsistent results.

Figure 9 depicts the final Gaussian approximations of the marginal posteriors $p(\theta_k|\mathbf{y})$, $k = 1, \dots, 8$, obtained by the pMCMC and the NPMC methods, in the CO and PO scenarios, for the average simulation runs represented as big circles and squares in Figure 6. We can observe that the NPMC method provides a significantly better approximation of the log-rate parameters in the CO scenario, where a larger amount of data is available, which is also clear from Figure 6 (*right*). However, the pMCMC on average performs similarly in both scenarios, due to the low efficiency of the pMCMC sampling scheme when the dimension of the problem (either K or N) increases.

In Table 10 the MSE of each parameter θ_k averaged over $P = 100$ independent simulation runs is shown, as obtained with the pMCMC and the NPMC schemes, for the CO and the PO experiments. In the CO case, NPMC provides homogeneous results for all parameters. On the contrary, in the PO case, some of the parameters (specially θ_5 and θ_6) are significantly poorly estimated, presenting a final MSE close to the initial value (which corresponds to the prior knowledge). The pMCMC scheme presents significantly higher MSE values than NPMC in both observation scenarios and for all parameters θ_k .

Figure 11 depicts the population posterior mean $\hat{\mathbf{x}} = E_{p(\mathbf{x}|\mathbf{y})}[\mathbf{x}]$ corresponding to the average simulation runs of the pMCMC and the NPMC methods in the PO scenario, represented as big squares in Figure 6. Again, the NPMC method provides more accurate estimates of the unobserved populations than the pMCMC method, specially for x_{RNA} . In the CO scenario both methods provide good approximations of the populations of all species.

6 Asymptotic convergence of NIS with approximate weights

6.1 Scope of the analysis

An analysis of the asymptotic effect of the transformation of the weights on the IS-based approximation of integrals w.r.t. a target probability distribution has already been addressed in (Koblents and Míguez, 2013b). In particular, the results in (Koblents and Míguez, 2013b) show that, as long as $\frac{M_T}{M} \rightarrow 0$, the distortion introduced by the *clipping* of the weights vanishes asymp-

totically and the approximation of integrals of bounded functions using IWs and using TIWs both converge to the same value almost surely (a.s.). However,

- the argument in (Koblents and Míguez, 2013b) is based on classical concentration-of-measure inequalities and, therefore, rates are only found for convergence in probability, and
- more importantly, the analysis relies on the ability to compute the non-normalized IWs exactly.

It is apparent from the algorithm description in Section 4 that, in the case of the SKM models of interest in this paper, the IWs can only be approximated (via particle filtering) and, therefore, the assumptions on which the theoretical results of (Koblents and Míguez, 2013b) rely are not satisfied. In this section, we improve on the analysis in (Koblents and Míguez, 2013b) by looking explicitly into the convergence of the approximations of integrals computed using approximate weights (both IWs and TIWs). We provide convergence rates for the L_p norms of the approximation errors and show that the approximate weights computed by a standard particle filter are “good enough” to ensure that these results hold.

6.2 Notation and basic assumptions

Let $\pi(\boldsymbol{\theta})$ be the pdf associated to the target probability distribution, let $q(\boldsymbol{\theta})$ be the importance function used to propose samples in an IS scheme (not necessarily normalized) and let $h(\boldsymbol{\theta}) \propto \pi(\boldsymbol{\theta})$ be a function proportional to π , with the proportionality constant independent of $\boldsymbol{\theta}$. The samples drawn from the distribution associated to q are denoted $\boldsymbol{\theta}^{(i)}$, $i = 1, \dots, M$, and their associated non-normalized IWs are $w^{(i)*} = h(\boldsymbol{\theta}^{(i)})/q(\boldsymbol{\theta}^{(i)})$, $i = 1, \dots, M$.

Let us define the weight function $g(\boldsymbol{\theta}) = h(\boldsymbol{\theta})/q(\boldsymbol{\theta})$ and, in particular, $g(\boldsymbol{\theta}^{(i)}) = w^{(i)*}$. The support of g is the same as the support of q , denoted $S \subseteq \mathbb{R}^K$. If we assume that both $q(\boldsymbol{\theta}) > 0$ and $\pi(\boldsymbol{\theta}) > 0$ for any $\boldsymbol{\theta} \in S$, then $g(\boldsymbol{\theta}) > 0$ for every $\boldsymbol{\theta} \in S$ as well. Also, trivially, $\pi \propto gq$, with the proportionality constant independent of $\boldsymbol{\theta}$. These assumptions are standard for classical IS.

Assume that the standard IWs can be computed exactly. In that case, the approximation π^M of the target probability measure can be written as

$$\pi^M(d\boldsymbol{\theta}) = \sum_{i=1}^M w^{(i)} \delta_{\boldsymbol{\theta}^{(i)}}(d\boldsymbol{\theta}),$$

where $w^{(i)} = \frac{g(\boldsymbol{\theta}^{(i)})}{\sum_{j=1}^M g(\boldsymbol{\theta}^{(j)})}$, $i = 1, \dots, M$.

Assume next that the weight function cannot be evaluated exactly but, instead, a sequence of approximations $g^J(\boldsymbol{\theta})$, $J \in \mathbb{N}$, exists for any point $\boldsymbol{\theta} \in \mathbf{S}$. We denote the random measure constructed from the approximate IWs as

$$\pi^{M,J}(d\boldsymbol{\theta}) = \sum_{i=1}^M w^{(i),J} \delta_{\boldsymbol{\theta}^{(i)}}(d\boldsymbol{\theta}),$$

where $w^{(i),J} = \frac{g^J(\boldsymbol{\theta}^{(i)})}{\sum_{j=1}^M g^J(\boldsymbol{\theta}^{(j)})}$, $i = 1, \dots, M$. Let us denote by φ^M the nonlinear transformation function used to compute non-normalized TIWs, i.e., $\bar{w}^{(i)*} = \varphi^M(w^{(i)*})$, $i = 1, \dots, M$, where $w^{(i)*}$ is the standard unnormalized IW associated to the sample $\boldsymbol{\theta}^{(i)}$. Then the weighted approximation of $\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ constructed according to the NIS scheme is

$$\bar{\pi}^{M,J}(d\boldsymbol{\theta}) = \sum_{i=1}^M \bar{w}^{(i),J} \delta_{\boldsymbol{\theta}^{(i)}}(d\boldsymbol{\theta}),$$

where $\bar{w}^{(i),J} = \frac{\varphi^M(g^J(\boldsymbol{\theta}^{(i)}))}{\sum_{j=1}^M \varphi^M(g^J(\boldsymbol{\theta}^{(j)}))}$, $i = 1, \dots, M$.

We make the following assumptions on the transformation function φ^M , the weight function g and its approximations $\{g^J : J \geq 1\}$.

A1 The nonlinear transformation φ^M of the weights is of a *clipping* class. In particular, given an index permutation i_1, \dots, i_M such that $w^{(i_1)*} \geq \dots \geq w^{(i_M)*}$, and a choice of the *clipping* parameter $M_T < M$, the transformation φ^M can be expressed as²

$$\varphi^M(w^{(i_k)*}) = \begin{cases} w^{(i_{M_T})*}, & \text{for } k = 1, \dots, M_T, \text{ and} \\ w^{(i_k)*}, & \text{for } k = M_T + 1, \dots, M. \end{cases}$$

A2 The weight function g has a finite upper bound and a positive lower bound. Specifically, there exists a real number $0 < a < \infty$ such that $a^{-1} \leq g(\boldsymbol{\theta}) \leq a$ for every $\boldsymbol{\theta} \in \mathbf{S}$.

A3 The same bounds of the weight function g hold for its approximations g^J , $J \geq 1$. To be specific, the inequalities $a^{-1} \leq g^J(\boldsymbol{\theta}) \leq a$ hold for every $\boldsymbol{\theta} \in \mathbf{S}$, any $J \geq 1$ and the same real number $0 < a < \infty$ as in A2.

A4 The approximation g^J of the weight function is possibly random and satisfies the inequality

$$\sup_{\boldsymbol{\theta} \in \mathbf{S}} |g(\boldsymbol{\theta}) - g^J(\boldsymbol{\theta})| \leq \frac{W_{g,\epsilon}}{J^{\frac{1}{2}-\epsilon}}$$

where $W_{g,\epsilon}$ is a positive a.s. finite random variable and $0 < \epsilon < \frac{1}{2}$ is an arbitrarily small constant, both independent of J .

² Note that φ^M is a function of both the complete weight set $\{w^{(j)*}\}_{j=1}^M$ and the index of the weight to be transformed, i.e., $\varphi^M : \{w^{(j)*}, j = 1, \dots, M\} \times \{1, \dots, M\} \rightarrow [1, +\infty)$.

Note that if the support set \mathbf{S} is compact then assumption A2 holds whenever $q > 0$ and $h > 0$ in \mathbf{S} . Otherwise, the proposal q has to be chosen so that it has heavier tails than π .

In the sequel we look into the approximation of integrals of the form $(f, \pi) = \int I_{\mathbf{S}}(\boldsymbol{\theta})f(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$, where $I_{\mathbf{S}}(\boldsymbol{\theta})$ is an indicator function³ and f is a bounded real function in the parameter space \mathbf{S} . We use $\|f\|_{\infty} = \sup_{\boldsymbol{\theta} \in \mathbf{S}} |f(\boldsymbol{\theta})| < \infty$ to denote the supremum norm of a bounded function. The set of bounded functions on \mathbf{S} is $B(\mathbf{S}) = \{f : \mathbf{S} \rightarrow \mathbb{R} : \|f\|_{\infty} < \infty\}$. The approximations of interest are

$$(f, \pi^{M,J}) = \sum_{i=1}^M f(\boldsymbol{\theta}^{(i)})w^{(i),J}, \quad \text{and} \\ (f, \bar{\pi}^{M,J}) = \sum_{i=1}^M f(\boldsymbol{\theta}^{(i)})\bar{w}^{(i),J}.$$

6.3 Convergence rates

The following basic Lemma establishes that both $(f, \pi^{M,J})$ and $(f, \bar{\pi}^{M,J})$ converge toward (f, π) a.s. and provides explicit rates for the absolute approximation errors.

Lemma 1 Assume that A1, A2, A3 and A4 hold,

$$J = J(M) \geq M \quad \text{and} \quad M_T \leq \sqrt{M}.$$

Then, there exist positive and a.s. finite random variables $W_{f,g,\epsilon}$ and $\bar{W}_{f,g,\epsilon}$, independent of M and J , such that

$$|(f, \pi^{M,J}) - (f, \pi)| \leq \frac{W_{f,g,\epsilon}}{M^{\frac{1}{2}-\epsilon}} \quad (5)$$

and

$$|(f, \bar{\pi}^{M,J}) - (f, \pi)| \leq \frac{\bar{W}_{f,g,\epsilon}}{M^{\frac{1}{2}-\epsilon}} \quad (6)$$

for every $f \in B(\mathbf{S})$, where $0 < \epsilon < \frac{1}{2}$ is an arbitrarily small constant independent of M and J . In particular

$$\lim_{M \rightarrow \infty} (f, \pi^{M,J}) = \lim_{M \rightarrow \infty} (f, \bar{\pi}^{M,J}) = (f, \pi) \text{ a.s.} \quad (7)$$

A proof is provided in Appendix B. Lemma 1 shows that we attain the usual Monte Carlo rate of convergence ($M^{-\frac{1}{2}+\epsilon}$) despite the approximation of the IWs and its subsequent *clipping* to compute TIWs. Note, however, that the random variables $W_{f,g,\epsilon}$ and $\bar{W}_{f,g,\epsilon}$ are not equal and, in general, $W_{f,g,\epsilon} \leq \bar{W}_{f,g,\epsilon}$.

³ Namely, $I_{\mathbf{S}}(\boldsymbol{\theta}) = 1$ if $\boldsymbol{\theta} \in \mathbf{S}$ and $I_{\mathbf{S}}(\boldsymbol{\theta}) = 0$ otherwise.

6.4 Approximate weights via particle filtering

In this section we introduce a more precise notation for the state-space model (compared to the argument-wise used in the previous sections), in order to perform the analysis with approximate weights. Assume we have a discrete-time state space Markov model with state process $\{\mathbf{X}_n\}_{n \geq 0}$ taking values on $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ and an observation process $\{\mathbf{Y}_n\}_{n \geq 0}$ taking values on $\mathcal{Y} \subseteq \mathbb{R}^{d_y}$. The prior distribution (probability measure) of the state is now denoted $\tau_0(d\mathbf{x})$ and the transition (Markov) kernel depends on a vector-valued random parameter $\boldsymbol{\Theta}$ that takes values on a compact set $S \subset \mathbb{R}^{d_\theta}$ and has prior distribution $\mu_0(d\boldsymbol{\theta})$ independent of \mathbf{X}_0 . In particular, the Markov kernel is now denoted $\tau_{n,\boldsymbol{\theta}}(d\mathbf{x}_n|\mathbf{x}_{n-1})$ and the conditional density of the observations is $u_n(\mathbf{y}_n|\mathbf{x}_n) > 0$. The latter also yields the likelihood of the signal \mathbf{x}_n , hence we often write, for conciseness, $u_n^{\mathbf{y}_n}(\mathbf{x}_n) \triangleq u_n(\mathbf{y}_n|\mathbf{x}_n)$.

At time n , the one-step-ahead predictive distribution of the state \mathbf{X}_n given fixed observations $\mathbf{Y}_{1:n-1} = \mathbf{y}_{1:n-1}$ and a parameter value $\boldsymbol{\Theta} = \boldsymbol{\theta}$ is denoted $\xi_{n,\boldsymbol{\theta}}$, specifically, for any Borel subset $A \subset \mathcal{X}$,

$$\xi_{n,\boldsymbol{\theta}}(A) = \mathbb{P}_n(\mathbf{X}_n \in A | \mathbf{Y}_{1:n-1} = \mathbf{y}_{1:n-1}, \boldsymbol{\Theta} = \boldsymbol{\theta})^4.$$

The filter measure at time n given observations $\mathbf{Y}_{1:n} = \mathbf{y}_{1:n}$ and parameter $\boldsymbol{\Theta} = \boldsymbol{\theta}$ is denoted $\phi_{n,\boldsymbol{\theta}}$, namely,

$$\phi_{n,\boldsymbol{\theta}}(A) = \mathbb{P}_n(\mathbf{X}_n \in A | \mathbf{Y}_{1:n} = \mathbf{y}_{1:n}, \boldsymbol{\Theta} = \boldsymbol{\theta}).$$

The predictive measure $\xi_{n,\boldsymbol{\theta}}$ can be expressed in terms of $\tau_{n,\boldsymbol{\theta}}$ and $\phi_{n-1,\boldsymbol{\theta}}$. Specifically, we write $\xi_{n,\boldsymbol{\theta}} = \tau_{n,\boldsymbol{\theta}}\phi_{n-1,\boldsymbol{\theta}}$, meaning that, for any integrable function $f : \mathcal{X} \rightarrow \mathbb{R}$,

$$\begin{aligned} (f, \xi_{n,\boldsymbol{\theta}}) &= \int \int f(\mathbf{x}) \tau_{n,\boldsymbol{\theta}}(d\mathbf{x}|\mathbf{x}') \phi_{n-1,\boldsymbol{\theta}}(d\mathbf{x}') \\ &= (f, \tau_{n,\boldsymbol{\theta}}\phi_{n-1,\boldsymbol{\theta}}). \end{aligned}$$

We also note that

$$(f, \xi_{n,\boldsymbol{\theta}}) = (\bar{f}_n, \phi_{n-1,\boldsymbol{\theta}}),$$

where $\bar{f}_n(\mathbf{x}') = \int f(\mathbf{x}) \tau_{n,\boldsymbol{\theta}}(d\mathbf{x}|\mathbf{x}')$. The filter measures $\phi_{n,\boldsymbol{\theta}}$ and $\phi_{n-1,\boldsymbol{\theta}}$ are related by the projective product

$$\phi_{n,\boldsymbol{\theta}} = u_n^{\mathbf{y}_n} \star \tau_{n,\boldsymbol{\theta}} \phi_{n-1,\boldsymbol{\theta}} = u_n^{\mathbf{y}_n} \star \xi_{n,\boldsymbol{\theta}},$$

defined as (Bain and Crisan, 2008)

$$(f, u_n^{\mathbf{y}_n} \star \xi_{n,\boldsymbol{\theta}}) \triangleq \frac{(f u_n^{\mathbf{y}_n}, \xi_{n,\boldsymbol{\theta}})}{(u_n^{\mathbf{y}_n}, \xi_{n,\boldsymbol{\theta}})}.$$

⁴ \mathbb{P}_n denotes the joint probability measure for the set of random variables $\{\mathbf{x}_k\}_{k \leq n} \cup \{\mathbf{y}_k\}_{k \leq n} \cup \{\boldsymbol{\Theta}\}$ on the measurable space $(\sigma(\mathbf{x}_{0:n}, \mathbf{y}_{1:n}, \boldsymbol{\Theta}), \mathcal{X}^{n+1} \times \mathcal{Y}^n \times S)$.

Let

$$\xi_{n,\boldsymbol{\theta}}^J(d\mathbf{x}) = \frac{1}{J} \sum_{j=1}^J \delta_{\mathbf{x}_{n,(j)}}(d\mathbf{x}) \text{ and}$$

$$\phi_{n,\boldsymbol{\theta}}^J(d\mathbf{x}) = \frac{1}{J} \sum_{j=1}^J \delta_{\bar{\mathbf{x}}_{n,(j)}}(d\mathbf{x})$$

be the approximations of $\xi_{n,\boldsymbol{\theta}}$ and $\phi_{n,\boldsymbol{\theta}}$ produced by a standard particle filter (Gordon et al, 1993) with J particles. We have the following theoretical guarantee for the convergence of $\xi_{n,\boldsymbol{\theta}}^J$ and $\phi_{n,\boldsymbol{\theta}}^J$.

Lemma 2 *Let N be a finite time horizon and let $\mathbf{Y}_{1:N} = \mathbf{y}_{1:N}$ be an arbitrary but fixed sequence of observations. Assume that, for every $n = 1, \dots, N$, $u_n^{\mathbf{y}_n} \in B(\mathcal{X})$, S is compact and*

$$\inf_{\boldsymbol{\theta} \in S} (u_n^{\mathbf{y}_n}, \xi_{n,\boldsymbol{\theta}}) > 0. \quad (8)$$

Then, for every $f \in B(\mathcal{X})$, every $p \geq 1$ and every $n = 0, 1, \dots, N$,

$$\sup_{\boldsymbol{\theta} \in S} \|(f, \xi_{n,\boldsymbol{\theta}}^J) - (f, \xi_{n,\boldsymbol{\theta}})\|_p \leq \frac{c_{1,n} \|f\|_\infty}{\sqrt{J}} \quad (9)$$

$$\sup_{\boldsymbol{\theta} \in S} \|(f, \phi_{n,\boldsymbol{\theta}}^J) - (f, \phi_{n,\boldsymbol{\theta}})\|_p \leq \frac{c_{2,n} \|f\|_\infty}{\sqrt{J}}, \quad (10)$$

where $c_{1,n}$ and $c_{2,n}$ are positive and finite constants independent of J and $\boldsymbol{\theta}$.

Proof. This is a straightforward consequence of (Crisan and Míguez, 2013, Lemma 2). \square

We denote the likelihood of the parameter realization $\boldsymbol{\theta}$ given the observations $\mathbf{Y}_{1:N} = \mathbf{y}_{1:N}$ as $\lambda_N(\boldsymbol{\theta})$, where

$$\lambda_N(\boldsymbol{\theta}) \triangleq \prod_{n=1}^N (u_n^{\mathbf{y}_n}, \xi_{n,\boldsymbol{\theta}})$$

(it is straightforward to show that $\lambda_N(\boldsymbol{\theta})$ yields the value of the joint pdf of $\mathbf{y}_1, \dots, \mathbf{y}_N$ conditional on $\boldsymbol{\theta}$). This likelihood can be naturally approximated via particle filtering as

$$\lambda_N^J(\boldsymbol{\theta}) \triangleq \prod_{n=1}^N (u_n^{\mathbf{y}_n}, \xi_{n,\boldsymbol{\theta}}^J)$$

and still guarantee that $\lambda_N^J \rightarrow \lambda_N$ a.s. with standard Monte Carlo rates. This is rigorously stated below.

Lemma 3 *Under the assumptions of Lemma 2 there exists a positive and a.s. finite random variable $W_{N,u,\epsilon}$ independent of J such that*

$$\sup_{\boldsymbol{\theta} \in S} |\lambda_N^J(\boldsymbol{\theta}) - \lambda_N(\boldsymbol{\theta})| \leq \frac{W_{N,u,\epsilon}}{J^{\frac{1}{2}-\epsilon}}, \quad (11)$$

where $0 < \epsilon < \frac{1}{2}$ is an arbitrarily small constant independent of J . In particular, the inequality (11) implies that $\lim_{J \rightarrow \infty} \lambda_N^J(\boldsymbol{\theta}) = \lambda_N(\boldsymbol{\theta})$ a.s. and uniformly over $\boldsymbol{\theta} \in S$.

Proof. See Appendix C. \square

6.5 Convergence of the NIS scheme with approximate weights

We can put the previous Lemmas together to prove convergence of the NIS scheme with approximate weights.

Assume that we use NIS to approximate the posterior measure of the parameter θ , namely

$$\pi(\theta)d\theta = \mathbb{P}_N(\Theta \in d\theta | \mathbf{Y}_{1:N} = \mathbf{y}_{1:N}). \quad (12)$$

It is straightforward to show that

$$\pi(\theta) \propto h(\theta) = \lambda_N(\theta)m_0(\theta),$$

where $m_0(\theta)$ is the density associated to the prior probability distribution of the parameter, μ_0 . If a proposal pdf q is used, the weight function becomes

$$g(\theta) = \frac{h(\theta)}{q(\theta)} = \frac{\lambda_N(\theta)m_0(\theta)}{q(\theta)}.$$

Since the likelihood $\lambda_N(\theta)$ cannot be computed in closed form we readily approximate it using a particle filter. This, in turn, yields the approximate weight function

$$g^J(\theta) = \frac{h^J(\theta)}{q(\theta)} = \frac{\lambda_N^J(\theta)m_0(\theta)}{q(\theta)}. \quad (13)$$

Let us apply a NIS scheme to approximate the target distribution in (12), where the weight function can be approximately evaluated using (13). The approximation of π with standard IWs is denoted $\pi^{M,J}$ and the approximation with TIWs is denoted $\bar{\pi}^{M,J}$. The observations $\mathbf{y}_{1:N}$ are arbitrary but fixed. Then we have the following result.

Theorem 1 *Assume that A1 holds, $J = J(M) \geq M$, $M_T \leq M$, $u_n^{\mathbf{y}_n} \in B(\mathcal{X})$ for every $n = 1, \dots, N$ and there exists a real constant $a > 0$ such that $\inf_{x \in \mathcal{X}} u_n^{\mathbf{y}_n} \geq \frac{1}{a}$ for every $n = 1, \dots, N$. If the inequalities*

$$\|m_0/q\|_\infty = \sup_{\theta \in \mathcal{S}} \frac{m_0(\theta)}{q(\theta)} < \infty, \quad (14)$$

$$\text{and } \inf_{\theta \in \mathcal{S}} \frac{m_0(\theta)}{q(\theta)} > 0$$

are satisfied, then, for every $f \in B(\mathcal{S})$, there exist positive random variables $W_{f,g,\epsilon}$ and $\bar{W}_{f,g,\epsilon}$, a.s. finite and independent of M and J , such that

$$|(f, \pi^{M,J}) - (f, \pi)| \leq \frac{W_{f,g,\epsilon}}{M^{\frac{1}{2}-\epsilon}}, \quad \text{and} \quad (15)$$

$$|(f, \bar{\pi}^{M,J}) - (f, \pi)| \leq \frac{\bar{W}_{f,g,\epsilon}}{M^{\frac{1}{2}-\epsilon}}, \quad (16)$$

where $0 < \epsilon < \frac{1}{2}$ is an arbitrarily small constant independent of M . The inequalities (15) and (16) imply

$$\lim_{M \rightarrow \infty} (f, \pi^{M,J}) = \lim_{M \rightarrow \infty} (f, \bar{\pi}^{M,J}) = (f, \pi) \quad \text{a.s.}$$

Proof. The absolute error in the approximation of the weight function is

$$|g(\theta) - g^J(\theta)| = \frac{m_0(\theta)}{q(\theta)} |\lambda_N^J(\theta) - \lambda_N(\theta)|. \quad (17)$$

However, from Lemma 3, we readily have⁵

$$\sup_{\theta \in \mathcal{S}} |\lambda_N^J(\theta) - \lambda_N(\theta)| \leq \frac{W_{N,u,\epsilon}}{J^{\frac{1}{2}-\epsilon}} \quad (18)$$

where $W_{N,u,\epsilon} > 0$ is a.s. finite and $0 < \epsilon < \frac{1}{2}$ is arbitrarily small, and both are independent of J (and M). Substituting (18) and (14) into (17) yields

$$\sup_{\theta \in \mathcal{S}} |g(\theta) - g^J(\theta)| \leq \frac{W_{N,u,\epsilon} \|m_0/q\|_\infty}{J^{\frac{1}{2}-\epsilon}}$$

and, as a consequence, the sequence of approximate weight functions g^J satisfies A4 with

$$W_{g,\epsilon} = \|m_0/q\|_\infty W_{N,u,\epsilon} > 0$$

a.s. finite.

Assumptions A2 and A3 are also satisfied. In particular, since $u_n^{\mathbf{y}_n} \in B(\mathcal{X})$ for every $n = 1, \dots, N$, it follows that

$$\prod_{n=1}^N (u_n^{\mathbf{y}_n}, \alpha) \leq \prod_{n=1}^N \|u_n^{\mathbf{y}_n}\|_\infty < \infty$$

for any probability measure on $(\mathcal{B}(\mathcal{X}), \mathcal{X})$ (where $\mathcal{B}(\mathcal{X})$ denotes the Borel σ -algebra of subsets of \mathcal{X}). In particular, $\prod_{n=1}^N \|u_n^{\mathbf{y}_n}\|_\infty$ is an upper bound for λ_N and λ_N^J . Moreover, since $\inf_{x \in \mathcal{X}} u_n^{\mathbf{y}_n} \geq a^{-1}$ for every $n = 1, \dots, N$ it follows that

$$\prod_{n=1}^N (u_n^{\mathbf{y}_n}, \alpha) \geq a^{-N} > 0$$

for any probability measure α on $(\mathcal{B}(\mathcal{X}), \mathcal{X})$. In particular, a^{-N} is a positive lower bound for both λ_N and λ_N^J . The factor m_0/q , independent of the approximation index J , has a positive lower bound and a finite upper bound by assumption.

Since A1–A4 are satisfied, we can apply Lemma 1, which yields (15) and (16) directly. \square

⁵ The assumptions of Theorem 1 imply the assumptions of Lemmas 2 and 3. In particular, $\inf_{x \in \mathcal{X}} u_n^{\mathbf{y}_n} \geq \frac{1}{a}$ implies $\inf_{\theta \in \mathcal{S}} (u_n^{\mathbf{y}_n}, \xi_{n,\theta}) > 0$.

7 Conclusion

We have addressed the problem of approximating posterior distributions of the parameters and the populations in stochastic kinetic models. We have applied a nonlinear population Monte Carlo (NPMC) method, which iteratively approximates the target distribution via an importance sampling scheme. The NPMC method resorts to a sequential Monte Carlo approximation of the posterior populations to evaluate the importance weights. Additionally, it performs nonlinear transformations to the weights to avoid degeneracy and the numerical problems typically arising in the proposal update of the PMC scheme in high dimensional problems. We provide an extended convergence analysis of the nonlinear importance sampling scheme, which takes into account the weight approximation.

We have compared the performance of the NPMC method to the well known particle Markov chain Monte Carlo (pMCMC) method, applied to the challenging prokaryotic autoregulatory model. Both methods have been applied in the exact simulation form, since the complexity of this model allows to do so. We show how the NPMC method outperforms the pMCMC method and requires only a moderate computational cost. Besides, the proposed method has a set of important features, common to all PMC schemes, as the sample independence, ease of parallelization, and compared to MCMC schemes, there is no need for convergence (burn-in) periods.

A Sequential Monte Carlo approximation of $p(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ and $p(\mathbf{y}|\boldsymbol{\theta})$

In this appendix we provide details on the approximation of the posterior $p(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ and the likelihood $p(\mathbf{y}|\boldsymbol{\theta})$. For a given vector of log-rate parameters $\boldsymbol{\theta}$, the following standard particle filter (see, e.g., (Doucet et al, 2001)) is run.

Initialization ($n = 0$):

Draw a collection of J samples $\{\mathbf{x}_0^{(j)}\}_{j=1}^J \sim p(\mathbf{x}_0)$.

Recursive step ($n = 1, \dots, N$):

1. Draw $\{\mathbf{x}_n^{(j)}\}_{j=1}^J \sim p(\mathbf{x}_n|\mathbf{x}_{n-1}^{(j)}, \boldsymbol{\theta})$ using the Gillespie algorithm (or a diffusion approximation).
2. Construct $\mathbf{x}_{1:n}^{(j)} = [\mathbf{x}_{1:n-1}^{(j)\top}, \mathbf{x}_n^{(j)\top}]^\top$.
3. Compute normalized IWs $\omega_n^{(j)*} = p(\mathbf{y}_n|\mathbf{x}_n^{(j)})$, $\omega_n^{(j)} = \omega_n^{(j)*} / \sum_{l=1}^J \omega_n^{(l)*}$, $j = 1, \dots, J$.
4. Resample J times with replacement from $\{\mathbf{x}_{1:n}^{(j)}\}_{j=1}^J$ according to the weights $\{\omega_n^{(j)}\}_{j=1}^J$ to yield $\{\tilde{\mathbf{x}}_{1:n}^{(j)}\}_{j=1}^J$.

An approximation of the posterior $p(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})d\mathbf{x}$ may be constructed from the final set of samples $\mathbf{x}_{1:N}^{(j)} = \mathbf{x}^{(j)}$ and

weights $\omega_N^{(j)}$ as the discrete random measure

$$\hat{p}^J(d\mathbf{x}|\boldsymbol{\theta}, \mathbf{y}) = \sum_{j=1}^J \omega_N^{(j)} \delta_{\mathbf{x}^{(j)}}(d\mathbf{x}).$$

The likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ can be approximated in turn as

$$\hat{p}^J(\mathbf{y}|\boldsymbol{\theta}) = \prod_{n=1}^N \frac{1}{J} \sum_{j=1}^J p(\mathbf{y}_n|\mathbf{x}_n^{(j)}).$$

In order to obtain a sample from the approximation $\hat{p}^J(d\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ in the pMCMC or the NPMC schemes, we just draw a sample out of the set $\{\mathbf{x}^{(j)}\}_{j=1}^J$ according to their IWs $\omega_N^{(j)}$.

B Proof of Lemma 1

We look into $(f, \pi^{M,J})$ first. Since

$$(f, \pi) = \frac{(fg, q)}{(g, q)} \text{ and } (f, \pi^{M,J}) = \frac{(fg^J, q^M)}{(g^J, q^M)}, \quad (19)$$

where $q^M = \frac{1}{M} \sum_{i=1}^M \delta_{\boldsymbol{\theta}^{(i)}}$, it is simple to show that

$$(f, \pi^{M,J}) - (f, \pi) = \frac{(fg^J, q^M) - (fg, q)}{(g, q)} + (f, \pi) \frac{(g, q) - (g^J, q^M)}{(g, q)}. \quad (20)$$

However, since $(g, q) = (1, h) = \int I_S(\boldsymbol{\theta})h(\boldsymbol{\theta})d\boldsymbol{\theta}$ and $(f, \pi) \leq \|f\|_\infty$, Eq. (20) readily yields

$$|(f, \pi^{M,J}) - (f, \pi)| \leq \frac{1}{(1, h)} |(fg^J, q^M) - (fg, q)| + \frac{\|f\|_\infty}{(1, h)} |(g, q) - (g^J, q^M)|, \quad (21)$$

and, therefore, the problem reduces to computing bounds for errors of the form $|(bg^J, q^M) - (bg, q)|$, where $b \in B(S)$.

Choose any $b \in B(S)$. A simple triangle inequality yields

$$|(bg^J, q^M) - (bg, q)| \leq |(bg^J, q^M) - (bg, q^M)| + |(bg, q^M) - (bg, q)|. \quad (22)$$

Since $q^M = \frac{1}{M} \sum_{i=1}^M \delta_{\boldsymbol{\theta}^{(i)}}$, for the second term on the right hand side of (22) we can write

$$\mathbb{E}[|(bg, q^M) - (bg, q)|^p] = \mathbb{E}\left[\left|\frac{1}{M} \sum_{i=1}^M Z^{(i)}\right|^p\right], \quad (23)$$

where the random variables

$$Z^{(i)} = b(\boldsymbol{\theta}^{(i)})g(\boldsymbol{\theta}^{(i)}) - (bg, q), \quad i = 1, \dots, M,$$

are i.i.d. with zero mean (since the $\boldsymbol{\theta}^{(i)}$'s are i.i.d. draws from q). Therefore, it is straightforward to show that

$$\mathbb{E}\left[\left|\frac{1}{M} \sum_{i=1}^M Z^{(i)}\right|^p\right] \leq \frac{\tilde{c}^p a^p \|b\|_\infty^p}{M^{\frac{p}{2}}}, \quad (24)$$

where \tilde{c} is a constant independent of M and q , and a is the uniform upper bound for the weight function g provided by

assumption A2, also independent of M . Combining (24) with (23) readily yields

$$\|(bg, q^M) - (bg, q)\|_p \leq \frac{\tilde{c}a\|b\|_\infty}{\sqrt{M}}. \quad (25)$$

The inequality (25) implies that there exists an a.s. finite random variable $U_\epsilon > 0$ such that

$$|(bg, q^M) - (bg, q)| \leq \frac{U_\epsilon}{M^{\frac{1}{2}-\epsilon}}, \quad (26)$$

where $0 < \epsilon < \frac{1}{2}$ is an arbitrarily small constant independent of M (see (Crisan and Míguez, 2011, Lemma 1)).

Expanding now the first term on the right hand side of (22) we find that

$$\begin{aligned} |(bg^J, q^M) - (bg, q^M)| &= \left| \frac{1}{M} \sum_{i=1}^M b(\theta^{(i)}) (g^J(\theta^{(i)}) - g(\theta^{(i)})) \right| \\ &\leq \frac{\|b\|_\infty}{M} \sum_{i=1}^M |g^J(\theta^{(i)}) - g(\theta^{(i)})|. \end{aligned} \quad (27)$$

However, by assumption A4, there exists an a.s. finite random variable $W_{g,\epsilon}$ such that

$$\sup_{\theta \in \mathcal{S}} |g^J(\theta) - g(\theta)| \leq \frac{W_{g,\epsilon}}{J^{\frac{1}{2}-\epsilon}}, \quad (28)$$

where $0 < \epsilon < \frac{1}{2}$ is an arbitrary small constant independent of J . Combining (28) with (27) yields

$$|(bg^J, q^M) - (bg, q^M)| \leq \frac{\|b\|_\infty W_{g,\epsilon}}{J^{\frac{1}{2}-\epsilon}}.$$

or, equivalently,

$$|(bg^J, q^M) - (bg, q^M)| \leq \frac{\|b\|_\infty W_{g,\epsilon}}{M^{\frac{1}{2}-\epsilon}}. \quad (29)$$

since we have assumed that $J = J(M) \geq M$.

Taking together (22), (26) and (29) we obtain

$$|(bg^J, q^M) - (bg, q)| \leq \frac{\|b\|_\infty W_{g,\epsilon} + U_\epsilon}{M^{\frac{1}{2}-\epsilon}} \quad (30)$$

and it is immediate to combine the inequality (21) with the bound in (30). If we choose $b = f$ in order to control the first term on the right hand side of (21), and $b = 1$ in order to control the second term, we readily find that

$$|(f, \pi^{M,J}) - (f, \pi)| \leq \frac{W_{f,g,\epsilon}}{M^{\frac{1}{2}-\epsilon}}, \quad (31)$$

where

$$W_{f,g,\epsilon} = \frac{1}{(1, h)} [(1 + \|f\|_\infty)W_{g,\epsilon} + 2U_\epsilon] > 0$$

is an a.s. finite random variable.

The proof for inequality (6) is simpler. A triangle inequality yields

$$|(f, \bar{\pi}^{M,J}) - (f, \pi)| \leq |(f, \bar{\pi}^{M,J}) - (f, \pi^{M,J})| + |(f, \pi^{M,J}) - (f, \pi)| \quad (32)$$

and (31) already provides an adequate bound for the second term on the right hand side of (32). For the first term on the right hand side, we note that

$$(f, \bar{\pi}^{M,J}) = \frac{(f[\varphi^M \circ g^J], q^M)}{(\varphi^M \circ g^J, q^M)}, \quad (33)$$

where \circ denotes composition, hence $(\varphi^M \circ g^J)(\theta) = \varphi^M(g^J(\theta))$. Taking together (33) and the expression for $(f, \pi^{M,J})$ in (19) yields, by the same argument leading to (21),

$$\begin{aligned} |(f, \bar{\pi}^{M,J}) - (f, \pi^{M,J})| &\leq \frac{|(f[\varphi^M \circ g^J], q^M) - (fg^J, q^M)|}{(\varphi^M \circ g^J, q^M)} \\ &\quad + \frac{\|f\|_\infty |(\varphi^M \circ g^J, q^M) - (g^J, q^M)|}{(\varphi^M \circ g^J, q^M)} \\ &\leq a|(f[\varphi^M \circ g^J], q^M) - (fg^J, q^M)| \\ &\quad + a\|f\|_\infty |(\varphi^M \circ g^J, q^M) - (g^J, q^M)|, \end{aligned} \quad (34)$$

where the second inequality follows from the definition of φ^M in A1 and the bound $g^J \geq a^{-1}$ in A3.

In order to use (34), we look into errors of the form $|(b[\varphi^M \circ g^J], q^M) - (bg^J, q^M)|$ for arbitrary $b \in B(\mathcal{S})$. This turns out relatively straightforward since, from the definition of φ^M in A1,

$$\begin{aligned} |(b[\varphi^M \circ g^J], q^M) - (bg^J, q^M)| &= \\ \left| \frac{1}{M} \sum_{r=1}^{M_T} b(\theta^{(i_r)}) [g^J(\theta^{(i_{M_T})}) - g^J(\theta^{(i_r)})] \right| &\leq 2a\|b\|_\infty \frac{M_T}{M}, \end{aligned} \quad (35)$$

where the inequality follows from using uniform bound $g^J \leq a$ in A3. We can plug (35) into (34) twice, first choosing $b = f$ and then $b = 1$, in order to control the two terms in the triangle inequality. As a result, we arrive at the *deterministic* bound

$$|(f, \bar{\pi}^{M,J}) - (f, \pi^{M,J})| \leq \frac{2a^2\|f\|_\infty M_T}{M} \leq \frac{2a^2\|f\|_\infty}{\sqrt{M}}, \quad (36)$$

where the second inequality follows from the assumption $M_T \leq \sqrt{M}$ in the statement of the Lemma.

Substituting (36) and (31) back into (32) yields

$$|(f, \bar{\pi}^{M,J}) - (f, \pi^{M,J})| \leq \frac{W_{f,g,\epsilon} + 2a^2\|f\|_\infty}{M^{\frac{1}{2}-\epsilon}}, \quad (37)$$

which reduces to the inequality (6) in the statement of the Lemma, with $\bar{W}_{f,g,\epsilon} = W_{f,g,\epsilon} + 2a^2\|f\|_\infty > 0$ an a.s. finite random variable. \square

C Proof of Lemma 3

It can be proved (Crisan and Míguez, 2013, Lemma 1) that for any $f \in B(\mathcal{X})$

$$\sup_{\theta \in \mathcal{S}} \|(f, \xi_{n,\theta}^J) - (f, \xi_{n,\theta})\|_p \leq \frac{c(f)}{\sqrt{J}}, \quad (38)$$

where $c(f)$ is a constant independent of θ and J . In particular, there exists an a.s. finite non negative random variable $U_{n,\theta,f,\epsilon}$, independent of J , such that

$$|(f, \xi_{n,\theta}^J) - (f, \xi_{n,\theta})| < \frac{U_{n,\theta,f,\epsilon}}{J^{\frac{1}{2}-\epsilon}}$$

for any constant $0 < \epsilon < \frac{1}{2}$ (see (Crisan and Míguez, 2011, Lemma 4.1)).

Note that, while the constant $c(f)$ in (38) is independent of θ , the random variable $U_{n,\theta,f,\epsilon}$ is not necessarily so. However, the inequality (38) holds for every $\theta \in \mathcal{S}$. Therefore $U_{n,\theta,f,\epsilon} \geq 0$ is a.s. finite for every $\theta \in \mathcal{S}$, hence

$$U_{n,f,\epsilon} := \sup_{\theta \in \mathcal{S}} U_{n,\theta,f,\epsilon} < \infty \quad \text{a.s.}$$

As a consequence, for any $f \in B(\mathcal{X})$,

$$\sup_{\theta \in \mathcal{S}} |(f, \xi_{n,\theta}^J) - (f, \xi_{n,\theta})| \leq \sup_{\theta \in \mathcal{S}} \frac{U_{n,f,\theta,\epsilon}}{J^{\frac{1}{2}-\epsilon}} \leq \frac{U_{n,f,\epsilon}}{J^{\frac{1}{2}-\epsilon}}, \quad (39)$$

where $U_{n,f,\epsilon} \geq 0$ is a.s. finite and independent of θ and J .

Now, given the record of observations $\mathbf{y}_{1:N}$ we need to find error rates for the likelihood of θ , namely for $\lambda_N(\theta) = \prod_{n=1}^N (u_n^{\mathbf{y}_n}, \xi_{n,\theta})$, where $u_n^{\mathbf{y}_n} \in B(\mathcal{X})$ and $\theta \in \mathcal{S}$. Using the inequality (39) we obtain

$$(u_n^{\mathbf{y}_n}, \xi_{n,\theta}) - \frac{U_{n,u,\epsilon}}{J^{\frac{1}{2}-\epsilon}} \leq (u_n^{\mathbf{y}_n}, \xi_{n,\theta}^J) \leq (u_n^{\mathbf{y}_n}, \xi_{n,\theta}) + \frac{U_{n,u,\epsilon}}{J^{\frac{1}{2}-\epsilon}} \quad (40)$$

a.s. for every $\theta \in \mathcal{S}$ (where the random variables $U_{n,u,\epsilon}$ is independent of θ and J , and a.s. finite) and, since $(u_n^{\mathbf{y}_n}, \xi_{n,\theta}^J) > 0$ by assumption, Eq. (40) readily yields

$$\begin{aligned} 0 \vee \prod_{n=1}^N \left[(u_n^{\mathbf{y}_n}, \xi_{n,\theta}) - \frac{U_{n,u,\epsilon}}{J^{\frac{1}{2}-\epsilon}} \right] &\leq \prod_{n=1}^N (u_n^{\mathbf{y}_n}, \xi_{n,\theta}^J) \\ &\leq \prod_{n=1}^N \left[(u_n^{\mathbf{y}_n}, \xi_{n,\theta}) + \frac{U_{n,u,\epsilon}}{J^{\frac{1}{2}-\epsilon}} \right], \end{aligned} \quad (41)$$

where $a \vee b$ denotes the maximum between a and b .

The term on the right hand side of (41) can be decomposed as

$$\begin{aligned} \prod_{n=1}^N \left[(u_n^{\mathbf{y}_n}, \xi_{n,\theta}) + \frac{U_{n,u,\epsilon}}{J^{\frac{1}{2}-\epsilon}} \right] &= \left(\prod_{n=1}^N (u_n^{\mathbf{y}_n}, \xi_{n,\theta}) \right) + \\ &\quad \sum_{\alpha \in A^N} \prod_{n=1}^N (u_n^{\mathbf{y}_n}, \xi_{n,\theta})^{\alpha_n} \times \left(\frac{U_{n,u,\epsilon}}{J^{\frac{1}{2}-\epsilon}} \right)^{1-\alpha_n}, \end{aligned}$$

where $\alpha = (\alpha_1, \dots, \alpha_N) \in \{0, 1\}^N$ is a multi-index of 0/1 entries and $A^N = \{0, 1\}^N \setminus (1, \dots, 1)$ is the set of all such multi-indices excluding $(1, \dots, 1)$. For every $\alpha \in A^N$, the factor $V_{N,u,\alpha_n,\epsilon} = \prod_{n=1}^N (u_n^{\mathbf{y}_n}, \xi_{n,\theta})^{\alpha_n} U_{n,u,\epsilon}^{1-\alpha_n}$ is a random variable and, since N is finite, $(u_n^{\mathbf{y}_n}, \xi_{n,\theta}) \leq \|u_n^{\mathbf{y}_n}\|_\infty < \infty$ and $U_{n,u,\epsilon} < \infty$ a.s., it turns out that

$$V_{N,u,\alpha_n,\epsilon} = \prod_{n=1}^N (u_n^{\mathbf{y}_n}, \xi_{n,\theta})^{\alpha_n} U_{n,u,\epsilon}^{1-\alpha_n} < \infty \quad \text{a.s.}$$

and, again, since $N < \infty$

$$V_{N,u,\epsilon} = \sum_{\alpha_n \in A^N} V_{N,u,\alpha_n,\epsilon} < \infty \quad \text{a.s.}$$

(a sum of a.s. finite random variables). Moreover, every $\alpha \in A^N$ contains at least one 0 entry, hence

$$\prod_{n=1}^N \left[(u_n^{\mathbf{y}_n}, \xi_{n,\theta}) + \frac{U_{n,u,\epsilon}}{J^{\frac{1}{2}-\epsilon}} \right] \leq \prod_{n=1}^N (u_n^{\mathbf{y}_n}, \xi_{n,\theta}) + \frac{V_{N,u,\epsilon}}{J^{\frac{1}{2}-\epsilon}}. \quad (42)$$

By a similar argument, there exists an a.s. finite random variable $\tilde{V}_{N,u,\epsilon}$ such that

$$\prod_{n=1}^N \left[(u_n^{\mathbf{y}_n}, \xi_{n,\theta}) - \frac{U_{n,u,\epsilon}}{J^{\frac{1}{2}-\epsilon}} \right] \geq \prod_{n=1}^N (u_n^{\mathbf{y}_n}, \xi_{n,\theta}) - \frac{\tilde{V}_{N,u,\epsilon}}{J^{\frac{1}{2}-\epsilon}}. \quad (43)$$

Combining (41), (42) and (43), we obtain

$$\begin{aligned} 0 \vee \left(\prod_{n=1}^N (u_n^{\mathbf{y}_n}, \xi_{n,\theta}) - \frac{\tilde{V}_{N,u,\epsilon}}{J^{\frac{1}{2}-\epsilon}} \right) &\leq \prod_{n=1}^N (u_n^{\mathbf{y}_n}, \xi_{n,\theta}^J) \\ &\leq \prod_{n=1}^N (u_n^{\mathbf{y}_n}, \xi_{n,\theta}) + \frac{V_{N,u,\epsilon}}{J^{\frac{1}{2}-\epsilon}}. \end{aligned} \quad (44)$$

Finally, if we introduce

$$W_{N,u,\epsilon} = V_{N,u,\epsilon} \vee \tilde{V}_{N,u,\epsilon} < \infty \quad \text{a.s.},$$

then (44) yields

$$\left| \prod_{n=1}^N (u_n^{\mathbf{y}_n}, \xi_{n,\theta}^J) - \prod_{n=1}^N (u_n^{\mathbf{y}_n}, \xi_{n,\theta}) \right| \leq \frac{W_{N,u,\epsilon}}{J^{\frac{1}{2}-\epsilon}},$$

where $0 \leq W_{N,u,\epsilon} < \infty$ a.s.

References

- Andrieu C, Doucet A, Holenstein R (2010) Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72(3):269–342
- Bain A, Crisan D (2008) *Fundamentals of stochastic filtering*, vol 60. Springer Verlag
- Bengtsson T, Bickel P, Li B (2008) Curse of dimensionality revisited: Collapse of particle filter in very large scale systems. *Probability and statistics: Essay in honour of David A Freedman* 2:316–334
- Boys RJ, Wilkinson DJ, Kirkwood TBL (2008) Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing* 18(2):125–135
- Cappé O, Guillin A, Marin JM, Robert CP (2004) Population Monte Carlo. *Computational and Graphical Statistics* 13(4):907–929
- Cappé O, Douc R, Guillin A, Marin JM, Robert CP (2008) Adaptive importance sampling in general mixture classes. *Statistics and Computing* 18(4):447–459
- Crisan D, Míguez J (2011) Particle approximation of the filtering density for state-space Markov models in discrete time. *arXiv preprint arXiv:11115866*
- Crisan D, Míguez J (2013) Nested particle filters for online parameter estimation in discrete-time state-space Markov models. *arXiv preprint arXiv:13081883*
- Doucet A, Godsill S, Andrieu C (2000) On sequential Monte Carlo Sampling methods for Bayesian filtering. *Statistics and Computing* 10(3):197–208
- Doucet A, De Freitas N, Gordon N (2001) *Sequential Monte Carlo methods in practice*. Springer Verlag
- Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* 81(25):2340–2361
- Golightly A, Wilkinson DJ (2005) Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics* 61(3):781–788
- Golightly A, Wilkinson DJ (2011) Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus* 1(6):807–820
- Gordon NJ, Salmond DJ, Smith AF (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In: *IEE Proceedings F (Radar and Signal Processing)*, IET, vol 140, pp 107–113

- Kilbinger Mea (2010) Bayesian model comparison in cosmology with population Monte Carlo. Royal astronomical society
- Kilbinger Mea (2012) CosmoPMC: Cosmology population Monte Carlo. arXiv preprint arXiv:11010950v3
- Koblenets E, Míguez J (2013a) A population Monte Carlo scheme for computational inference in high dimensional spaces. ICASSP
- Koblenets E, Míguez J (2013b) A population Monte Carlo scheme with transformed weights and its application to stochastic kinetic models. *Statistics and Computing* pp 1–19, DOI 10.1007/s11222-013-9440-2, URL <http://dx.doi.org/10.1007/s11222-013-9440-2>
- Koblenets E, Míguez J (2013c) Robust mixture population Monte Carlo scheme with adaptation of the number of components. EUSIPCO
- Lewis A, Bridle S (2002) Cosmological parameters from cmb and other data: a Monte Carlo approach. *Phys Rev D* 66:103,511, astro-ph/0205436
- Milner P, Gillespie C, Wilkinson D (2013) Moment closure based parameter inference of stochastic kinetic models. *Statistics and Computing* pp 1–9
- Robert CP, Casella G (2004) *Monte Carlo Statistical Methods*. Springer
- Volterra V (1926) Fluctuations in the abundance of a species considered mathematically. *Nature* 118:558–560
- Wilkinson D (2011a) Parameter inference for stochastic kinetic models of bacterial gene regulation: A Bayesian approach to systems biology. (with discussion), in *Bayesian Statistics* 9
- Wilkinson D (2011b) *Stochastic modelling for systems biology*, vol 44. CRC press
- Wraith Dea (2009) Estimation of cosmological parameters using adaptive importance sampling. arXiv preprint arXiv:09030837v1